BG

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification⁷: C12N 15/16, 15/19, C07K 14/52, 14/575, 16/24, 16/26, C12Q 1/68, A61K 38/19, 38/22, G01N 33/53

(21) International Application Number: PCT/US01/18476

(22) International Filing Date: 6 June 2001 (06.06.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/210,233   6 June 2000 (06.06.2000)   US
60/213,465   23 June 2000 (23.06.2000)   US
60/249,019   14 November 2000 (14.11.2000)   US

(71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and
(75) Inventors/Applicants (for US only): LAL, Preeti [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). HE, Ann [CN/US]; 4601 Catalina Drive, San Jose, CA 95129 (US). NGUYEN, Danniel, B. [US/US]; 1403 Ridgewood Drive, San Jose, CA 95118 (US). WALIA, Narinder [US/US]; 890 Davis Street #205, San Leandro, CA 94577 (US). GANDHI, Ameena, R. [US/US]; 837 Roble Avenue, #1, Menlo Park, CA 94025 (US). AZIMZAI, Yalda [US/US]; 5518 Boulder Canyon Drive, Castro Valley, CA 94552 (US). BANDMAN, Olga [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). TANG, Y., Tom [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). LU, Yan [US/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). DUGGAN, Brendan, M. [AU/US]; 243 Buena Vista, #306, Sunnyvale, CA 94086 (US). LEE, Sally [US/US]; 825 East Evelyn, #425, Sunnyvale, CA 94086 (US). HAFALIA, April [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). POLICKY, Jennifer, L. [US/US]; 1511 Jarvis Court, San Jose, CA 95118 (US).

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: EXTRACELLULAR MESSENGERS

(57) Abstract: The invention provides human extracellular messengers (XMES) and polynucleotides which identify and encode XMES. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating or preventing disorders associated with aberrant expression of XMES.

WO 01/94587 A2

EXTRACELLULAR MESSENGERS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of extracellular messengers and
5   to the use of these sequences in the diagnosis, treatment, and prevention of neurological disorders,
autoimmune/inflammatory disorders, developmental disorders, endocrine disorders, and cell
proliferative disorders including cancer, and in the assessment of the effects of exogenous compounds
on the expression of nucleic acid and amino acid sequences of extracellular messengers.

BACKGROUND OF THE INVENTION
10

Intercellular communication is essential for the growth and survival of multicellular
organisms, and in particular, for the function of the endocrine, nervous, and immune systems. In
addition, intercellular communication is critical for developmental processes such as tissue
construction and organogenesis, in which cell proliferation, cell differentiation, and morphogenesis
15   must be spatially and temporally regulated in a precise and coordinated manner. Cells communicate
with one another through the secretion and uptake of diverse types of signaling molecules such as
hormones, growth factors, neuropeptides, and cytokines.

Hormones

Hormones are signaling molecules that coordinately regulate basic physiological processes
20   from embryogenesis throughout adulthood. These processes include metabolism, respiration,
reproduction, excretion, fetal tissue differentiation and organogenesis, growth and development,
homeostasis, and the stress response. Hormonal secretions and the nervous system are tightly
integrated and interdependent. Hormones are secreted by endocrine glands, primarily the hypothalamus
and pituitary, the thyroid and parathyroid, the pancreas, the adrenal glands, and the ovaries and testes.

25   The secretion of hormones into the circulation is tightly controlled. Hormones are often
secreted in diurnal, pulsatile, and cyclic patterns. Hormone secretion is regulated by perturbations in
blood biochemistry, by other upstream-acting hormones, by neural impulses, and by negative feedback
loops. Blood hormone concentrations are constantly monitored and adjusted to maintain optimal,
steady-state levels. Once secreted, hormones act only on those target cells that express specific
30   receptors.

Most disorders of the endocrine system are caused by either hyposecretion or hypersecretion of
hormones. Hyposecretion often occurs when a hormone's gland of origin is damaged or otherwise
impaired. Hypersecretion often results from the proliferation of tumors derived from hormone-secreting
cells. Inappropriate hormone levels may also be caused by defects in regulatory feedback loops or in

the processing of hormone precursors. Endocrine malfunction may also occur when the target cell fails to respond to the hormone.

Hormones can be classified biochemically as polypeptides, steroids, eicosanoids, or amines. Polypeptide hormones, which include diverse hormones such as insulin and growth hormone, vary in
5   size and function and are often synthesized as inactive precursors that are processed intracellularly into mature, active forms. Amine hormones, which include epinephrine and dopamine, are amino acid derivatives that function in neuroendocrine signaling. Steroid hormones, which include the cholesterol-derived hormones estrogen and testosterone, function in sexual development and reproduction. Eicosanoid hormones, which include prostaglandins and prostacyclins, are fatty acid derivatives that
10   function in a variety of processes. Most polypeptide hormones and some amine hormones are soluble in the circulation where they are highly susceptible to proteolytic degradation within seconds after their secretion. Steroid hormones and eicosanoid hormones are insoluble and must be transported in the circulation by carrier proteins. The following discussion will focus primarily on polypeptide hormones.

Hormones secreted by the hypothalamus and pituitary gland play a critical role in endocrine
15   function by coordinately regulating hormonal secretions from other endocrine glands in response to neural signals. Hypothalamic hormones include thyrotropin-releasing hormone, gonadotropin-releasing hormone, somatostatin, growth-hormone releasing factor, corticotropin-releasing hormone, substance P, dopamine, and prolactin-releasing hormone. These hormones directly regulate the secretion of hormones from the anterior lobe of the pituitary. Hormones secreted by the anterior pituitary include
20   adrenocorticotropic hormone (ACTH), melanocyte-stimulating hormone, somatotropic hormones such as growth hormone and prolactin, glycoprotein hormones such as thyroid-stimulating hormone, luteinizing hormone (LH), and follicle-stimulating hormone (FSH), β-lipotropin, and β-endorphins. These hormones regulate hormonal secretions from the thyroid, pancreas, and adrenal glands, and act directly on the reproductive organs to stimulate ovulation and spermatogenesis. The posterior
25   pituitary synthesizes and secretes antidiuretic hormone (ADH, vasopressin) and oxytocin.

Disorders of the hypothalamus and pituitary often result from lesions such as primary brain tumors, adenomas, infarction associated with pregnancy, hypophysectomy, aneurysms, vascular malformations, thrombosis, infections, immunological disorders, and complications due to head trauma. Such disorders have profound effects on the function of other endocrine glands. Disorders
30   associated with hypopituitarism include hypogonadism, Sheehan syndrome, diabetes insipidus, Kallman's disease, Hand-Schuller-Christian disease, Letterer-Siwe disease, sarcoidosis, empty sella syndrome, and dwarfism. Disorders associated with hyperpituitarism include acromegaly, giantism, and syndrome of inappropriate ADH secretion (SIADH), often caused by benign adenomas.

Hormones secreted by the thyroid and parathyroid primarily control metabolic rates and the

regulation of serum calcium levels, respectively. Thyroid hormones include calcitonin, somatostatin, and thyroid hormone. The parathyroid secretes parathyroid hormone. Disorders associated with hypothyroidism include goiter, myxedema, acute thyroiditis associated with bacterial infection, subacute thyroiditis associated with viral infection, autoimmune thyroiditis (Hashimoto's disease),

5   and cretinism. Disorders associated with hyperthyroidism include thyrotoxicosis and its various forms, Grave's disease, pretibial myxedema, toxic multinodular goiter, thyroid carcinoma, and Plummer's disease. Disorders associated with hyperparathyroidism include Conn disease (chronic hypercalemia) leading to bone resorption and parathyroid hyperplasia.

       Hormones secreted by the pancreas regulate blood glucose levels by modulating the rates of

10   carbohydrate, fat, and protein metabolism. Pancreatic hormones include insulin, glucagon, amylin, γ-aminobutyric acid, gastrin, somatostatin, and pancreatic polypeptide. The principal disorder associated with pancreatic dysfunction is diabetes mellitus caused by insufficient insulin activity. Diabetes mellitus is generally classified as either Type I (insulin-dependent, juvenile diabetes) or Type II (non-insulin-dependent, adult diabetes). The treatment of both forms by insulin replacement

15   therapy is well known. Diabetes mellitus often leads to acute complications such as hypoglycemia (insulin shock), coma, diabetic ketoacidosis, lactic acidosis, and chronic complications leading to disorders of the eye, kidney, skin, bone, joint, cardiovascular system, nervous system, and to decreased resistance to infection.

       The anatomy, physiology, and diseases related to hormonal function are reviewed in

20   McCance, K. L. and Huether, S. E. (1994) Pathophysiology: The Biological Basis for Disease in Adults and Children, Mosby-Year Book, Inc., St. Louis, MO; Greenspan, F. S. and Baxter, J. D. (1994) Basic and Clinical Endocrinology, Appleton and Lange, East Norwalk, CT.

Growth Factors

       Growth factors are secreted proteins that mediate intercellular communication. Unlike

25   hormones, which travel great distances via the circulatory system, most growth factors are primarily local mediators that act on neighboring cells. Most growth factors contain a hydrophobic N-terminal signal peptide sequence which directs the growth factor into the secretory pathway. Most growth factors also undergo post-translational modifications within the secretory pathway. These modifications can include proteolysis, glycosylation, phosphorylation, and intramolecular disulfide

30   bond formation. Once secreted, growth factors bind to specific receptors on the surfaces of neighboring target cells, and the bound receptors trigger intracellular signal transduction pathways. These signal transduction pathways elicit specific cellular responses in the target cells. These responses can include the modulation of gene expression and the stimulation or inhibition of cell division, cell differentiation, and cell motility.

35   Growth factors fall into at least two broad and overlapping classes. The broadest class

3

includes the large polypeptide growth factors, which are wide-ranging in their effects. These factors include epidermal growth factor (EGF), fibroblast growth factor (FGF), transforming growth factor-β (TGF-β), insulin-like growth factor (IGF), nerve growth factor (NGF), and platelet-derived growth factor (PDGF), each defining a family of numerous related factors. The large polypeptide growth

5   factors, with the exception of NGF, act as mitogens on diverse cell types to stimulate wound healing, bone synthesis and remodeling, extracellular matrix synthesis, and proliferation of epithelial, epidermal, and connective tissues. Members of the TGF-β, EGF, and FGF families also function as inductive signals in the differentiation of embryonic tissue. NGF functions specifically as a neurotrophic factor, promoting neuronal growth and differentiation.

10   Another class of growth factors includes the hematopoietic growth factors, which are narrow in their target specificity. These factors stimulate the proliferation and differentiation of blood cells such as B-lymphocytes, T-lymphocytes, erythrocytes, platelets, eosinophils, basophils, neutrophils, macrophages, and their stem cell precursors. These factors include the colony-stimulating factors (G-CSF, M-CSF, GM-CSF, and CSF1-3), erythropoietin, and the cytokines. The cytokines are

15   specialized hematopoietic factors secreted by cells of the immune system and are discussed in detail below.

        Growth factors play critical roles in neoplastic transformation of cells in vitro and in tumor progression in vivo. Overexpression of the large polypeptide growth factors promotes the proliferation and transformation of cells in culture. Inappropriate expression of these growth factors

20   by tumor cells in vivo may contribute to tumor vascularization and metastasis. Inappropriate activity of hematopoietic growth factors can result in anemias, leukemias, and lymphomas. Moreover, growth factors are both structurally and functionally related to oncoproteins, the potentially cancer-causing products of proto-oncogenes. Certain FGF and PDGF family members are themselves homologous to oncoproteins, whereas receptors for some members of the EGF, NGF, and FGF families are encoded

25   by proto-oncogenes. Growth factors also affect the transcriptional regulation of both proto-oncogenes and oncosuppressor genes. (Pimentel, E. (1994) Handbook of Growth Factors, CRC Press, Ann Arbor, MI; McKay, I. and Leigh, I., eds. (1993) Growth Factors: A Practical Approach, Oxford University Press, New York, NY; Habenicht, A., ed. (1990) Growth Factors, Differentiation Factors, and Cytokines, Springer-Verlag, New York, NY.)

30   In addition, some of the large polypeptide growth factors play crucial roles in the induction of the primordial germ layers in the developing embryo. This induction ultimately results in the formation of the embryonic mesoderm, ectoderm, and endoderm which in turn provide the framework for the entire adult body plan. Disruption of this inductive process would be catastrophic to embryonic development.

Small Peptide Factors - Neuropeptides and Vasomediators

Neuropeptides and vasomediators (NP/VM) comprise a family of small peptide factors, typically of 20 amino acids or less. These factors generally function in neuronal excitation and inhibition of vasoconstriction/vasodilation, muscle contraction, and hormonal secretions from the

5   brain and other endocrine tissues. Included in this family are neuropeptides and neuropeptide hormones such as bombesin, neuropeptide Y, neurotensin, neuromedin N, melanocortins, opioids, galanin, somatostatin, tachykinins, urotensin II and related peptides involved in smooth muscle stimulation, vasopressin, vasoactive intestinal peptide, and circulatory system-borne signaling molecules such as angiotensin, complement, calcitonin, endothelins, formyl-methionyl peptides,

10   glucagon, cholecystokinin, gastrin, and many of the peptide hormones discussed above. NP/VMs can transduce signals directly, modulate the activity or release of other neurotransmitters and hormones, and act as catalytic enzymes in signaling cascades. The effects of NP/VMs range from extremely brief to long-lasting. (Reviewed in Martin, C. R. et al. (1985) Endocrine Physiology, Oxford University Press, New York, NY, pp. 57-62.)

15   Cytokines

Cytokines comprise a family of signaling molecules that modulate the immune system and the inflammatory response. Cytokines are usually secreted by leukocytes, or white blood cells, in response to injury or infection. Cytokines function as growth and differentiation factors that act primarily on cells of the immune system such as B- and T-lymphocytes, monocytes, macrophages,

20   and granulocytes. Like other signaling molecules, cytokines bind to specific plasma membrane receptors and trigger intracellular signal transduction pathways which alter gene expression patterns. There is considerable potential for the use of cytokines in the treatment of inflammation and immune system disorders.

Cytokine structure and function have been extensively characterized in vitro. Most cytokines

25   are small polypeptides of about 30 kilodaltons or less. Over 50 cytokines have been identified from human and rodent sources. Examples of cytokine subfamilies include the interferons (IFN-α, -β, and -γ), the interleukins (IL1-IL13), the tumor necrosis factors (TNF-α and -β), and the chemokines. Many cytokines have been produced using recombinant DNA techniques, and the activities of individual cytokines have been determined in vitro. These activities include regulation of leukocyte

30   proliferation, differentiation, and motility.

The activity of an individual cytokine in vitro may not reflect the full scope of that cytokine's activity in vivo. Cytokines are not expressed individually in vivo but are instead expressed in combination with a multitude of other cytokines when the organism is challenged with a stimulus. Together, these cytokines collectively modulate the immune response in a manner appropriate for that

35   particular stimulus. Therefore, the physiological activity of a cytokine is determined by the stimulus

5

itself and by complex interactive networks among co-expressed cytokines which may demonstrate both synergistic and antagonistic relationships.

Chemokines comprise a cytokine subfamily with over 30 members. (Reviewed in Wells, T. N. C. and Peitsch, M. C. (1997) J. Leukoc. Biol. 61:545-550.) Chemokines were initially identified

5   as chemotactic proteins that recruit monocytes and macrophages to sites of inflammation. Recent evidence indicates that chemokines may also play key roles in hematopoiesis and HIV-1 infection. Chemokines are small proteins which range from about 6-15 kilodaltons in molecular weight. Chemokines are further classified as C, CC, CXC, or CX_3C based on the number and position of critical cysteine residues. The CC chemokines, for example, each contain a conserved motif

10   consisting of two consecutive cysteines followed by two additional cysteines which occur downstream at 24- and 16-residue intervals, respectively (ExPASy PROSITE database, documents PS00472 and PDOC00434). The presence and spacing of these four cysteine residues are highly conserved, whereas the intervening residues diverge significantly. However, a conserved tyrosine located about 15 residues downstream of the cysteine doublet seems to be important for chemotactic

15   activity. Most of the human genes encoding CC chemokines are clustered on chromosome 17, although there are a few examples of CC chemokine genes that map elsewhere. Other chemokines include lymphotactin (C chemokine); macrophage chemotactic and activating factor (MCAF/MCP-1; CC chemokine); platelet factor 4 and IL-8 (CXC chemokines); and fractalkine and neurotractin (CX_3C chemokines). (Reviewed in Luster, A. D. (1998) N. Engl. J. Med. 338:436-445.)

20   The discovery of new extracellular messengers and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of neurological disorders, autoimmune/inflammatory disorders, developmental disorders, endocrine disorders, and cell proliferative disorders including cancer, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of

25   extracellular messengers.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, extracellular messengers, referred to collectively as "XMES" and individually as "XMES-1," "XMES-2," "XMES-3," "XMES-4," "XMES-5,"

30   "XMES-6," "XMES-7," "XMES-8," and "XMES-9." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, c) a biologically active fragment of a polypeptide having an amino acid

6

sequence selected from the group consisting of SEQ ID NO:1-9, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-9.

5   The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the

10   group consisting of SEQ ID NO:1-9, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-9. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:10-18.

          Additionally, the invention provides a recombinant polynucleotide comprising a promoter

15   sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of

20   SEQ ID NO:1-9, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

          The invention also provides a method for producing a polypeptide selected from the group

25   consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, and d) an immunogenic fragment of a polypeptide having an amino acid

30   sequence selected from the group consisting of SEQ ID NO:1-9. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

          Additionally, the invention provides an isolated antibody which specifically binds to a

7

polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, c) a biologically active fragment of a polypeptide having an amino acid

5   sequence selected from the group consisting of SEQ ID NO:1-9, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:10-18, b) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90%

10   identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:10-18, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample,

15   said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:10-18, b) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:10-18, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the

20   polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of

25   said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID

30   NO:10-18, b) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:10-18, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b)

8

detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from
5   the group consisting of SEQ ID NO:1-9, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, and a pharmaceutically
10   acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-9. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional XMES, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an
15   agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, and
20   d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of
25   treating a disease or condition associated with decreased expression of functional XMES, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, b) a naturally occurring
30   polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9. The method comprises a) exposing a sample comprising the

polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional XMES, comprising

5   administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the

10 group consisting of SEQ ID NO:1-9, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a

15 compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence

20 selected from the group consisting of SEQ ID NO:1-9, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the

25 activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in

30 altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:10-18, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said

method comprising a) treating a biological sample containing nucleic acids with the test compound;
b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20
contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide
comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:10-18, ii) a
5   naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a
polynucleotide sequence selected from the group consisting of SEQ ID NO:10-18, iii) a
polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the
polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions
whereby a specific hybridization complex is formed between said probe and a target polynucleotide in
10   the biological sample, said target polynucleotide selected from the group consisting of i) a
polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID
NO:10-18, ii) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90%
identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:10-18, iii) a
polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the
15   polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide
comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above;
c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization
complex in the treated biological sample with the amount of hybridization complex in an untreated
biological sample, wherein a difference in the amount of hybridization complex in the treated
20   biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide
sequences of the present invention.
25   Table 2 shows the GenBank identification number and annotation of the nearest GenBank
homolog for polypeptides of the invention. The probability score for the match between each
polypeptide and its GenBank homolog is also shown.
Table 3 shows structural features of polypeptide sequences of the invention, including predicted
motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of
30   the polypeptides.
Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble
polynucleotide sequences of the invention, along with selected fragments of the polynucleotide
sequences.
Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

5

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing

10   particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a

15   reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to

20   practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

25   **DEFINITIONS**

"XMES" refers to the amino acid sequences of substantially purified XMES obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of

30   XMES. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of XMES either by directly interacting with XMES or by acting on components of the biological pathway in which XMES participates.

An "allelic variant" is an alternative form of the gene encoding XMES. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in

polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in

5   a given sequence.

"Altered" nucleic acid sequences encoding XMES include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as XMES or a polypeptide with at least one functional characteristic of XMES. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of

10   the polynucleotide encoding XMES, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding XMES. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent XMES. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility,

15   hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of XMES is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains

20   having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring

25   protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

30   The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of XMES. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of XMES either by directly interacting with XMES or by acting on components of the biological pathway in which XMES participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')$_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind XMES polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used

5   to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that

10   makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

15   The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having

20   modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the

25   designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic XMES, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific

30   antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a

given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding XMES or fragments of XMES may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be

5   associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems,

10   Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

15   "Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

20

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

40

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide

15

backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

A "fragment" is a unique portion of XMES or the polynucleotide encoding XMES which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:10-18 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:10-18, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:10-18 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:10-18 from related polynucleotide sequences. The precise length of a fragment of SEQ

16

ID NO:10-18 and the region of SEQ ID NO:10-18 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-9 is encoded by a fragment of SEQ ID NO:10-18. A fragment of SEQ ID NO:1-9 comprises a region of unique amino acid sequence that specifically identifies SEQ

5   ID NO:1-9. For example, a fragment of SEQ ID NO:1-9 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-9. The precise length of a fragment of SEQ ID NO:1-9 and the region of SEQ ID NO:1-9 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

10   A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

15   The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

20   Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191.

25   For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is

30   provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other

17

polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST

5   programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

10   *Penalty for mismatch: -2*

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

15   *Filter: on*

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such

20   lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in

25   a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment

30   methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment

program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity"

5   between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

10   *Open Gap: 11 and Extension Gap: 1 penalties*

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

15   Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length

20   supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

25   The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific

30   hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for

annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v)

5   SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the

10   target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention

15   include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as

20   formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

25   The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells

30   or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of

various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of XMES which is capable of eliciting an immune response when introduced into a living organism, for example, a
5   mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of XMES which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other
10   chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of XMES. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of XMES.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide,
15   polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably
20   linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of
25   amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an XMES may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the
30   art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of XMES.

"Probe" refers to nucleic acid sequences encoding XMES, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical

labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid

5   sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may

10   be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular

15   Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

20   Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU

25   primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer

30   binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that

hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing

5   primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence.

10 This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence.

15 Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated

20 regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent,

25 chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose

30 instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing XMES, nucleic acids encoding XMES, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For

5   example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free,

10   preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters,

15   chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

20   "Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection,

25   electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to

30   animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in

24

vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection,

5   transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the

10   nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as

15   defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting

20   polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

25   A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at

30   least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human extracellular messengers (XMES), the

polynucleotides encoding XMES, and the use of these compositions for the diagnosis, treatment, or
prevention of neurological disorders, autoimmune/inflammatory disorders, developmental disorders,
endocrine disorders, and cell proliferative disorders including cancer.

   Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide
5 sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a
single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted
by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte
polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is
denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an
10 Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

   Table 2 shows sequences with homology to the polypeptides of the invention as identified by
BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the
polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte
polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3
15 shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog.
Column 4 shows the probability score for the match between each polypeptide and its GenBank
homolog. Column 5 shows the annotation of the GenBank homolog along with relevant citations where
applicable, all of which are expressly incorporated by reference herein.

   Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2
20 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte
polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3
shows the number of amino acid residues in each polypeptide. Column 4 shows potential
phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS
program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI).
25 Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7
shows analytical methods for protein structure/function analysis and in some cases, searchable
databases to which the analytical methods were applied.

   Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these
properties establish that the claimed polypeptides are extracellular messengers. For example, SEQ ID
30 NO:2 is 96% identical, over 1210 amino acids, to rat neurexin III-alpha (GenBank ID g394600) as
determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST
probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence
alignment by chance. Based on this analysis, SEQ ID NO:2 is a human neurexin. In an alternative
example, SEQ ID NO:1 is 39% identical over 240 amino acids to mac25 (g3721617) (a protein which

binds the class of proteins known as insulin-like growth factors) as determined by the Basic Local
Alignment Search Tool (BLAST). The probability score is 3.3e-42 (see Table 2). In an additional
alternative example, SEQ ID NO:3 is 39% identical, to human fibroblast growth factor 19 (GenBank
ID g4514718) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The

5   BLAST probability score is 1.5e-22, which indicates the probability of obtaining the observed
polypeptide sequence alignment by chance. SEQ ID NO:3 also contains a fibroblast growth factor
domain as determined by searching for statistically significant matches in the hidden Markov model
(HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from
BLIMPS and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:3 is a

10   fibroblast growth factor. In another alternative example, SEQ ID NO:6 is 81% identical to Bos taurus
cartilage-derived morphogenetic protein (GenBank ID g632490) as determined by the Basic Local
Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 7.5e-185, which
indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID
NO:6 also contains a TGF-beta propeptide domain and a TGF-beta-like domain as determined by

15   searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM
database of conserved protein family domains. (See Table 3.) Data from BLIMPS analysis provides
further corroborative evidence that SEQ ID NO:6 is a TGF-beta family protein. SEQ ID NO:4, SEQ
ID NO:5, SEQ ID NO:7, SEQ ID NO:8 and SEQ ID NO:9 were analyzed and annotated in a similar
manner. The algorithms and parameters for the analysis of SEQ ID NO:1-9 are described in Table 7.

20   As shown in Table 4, the full length polynucleotide sequences of the present invention were
assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any
combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence
identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide
consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention.

25   Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of
the polynucleotide sequences which are useful, for example, in hybridization or amplification
technologies that identify SEQ ID NO:10-18 or that distinguish between SEQ ID NO:10-18 and
related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA
sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages

30   comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length
polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5′)
and stop (3′) positions of the cDNA and/or genomic sequences in column 5 relative to their respective
full length sequences.

          The identification numbers in Column 5 of Table 4 may refer specifically, for example, to

Incyte cDNAs along with their corresponding cDNA libraries. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 70656442V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., g3153915) which contributed to the assembly of the full length polynucleotide sequences. Alternatively, the identification

5   numbers in column 5 may refer to coding regions predicted by Genscan analysis of genomic DNA. For example, GNN.g7768040_000033_002 is the identification number of a Genscan-predicted coding sequence, with g7768040 being the GenBank identification number of the sequence to which Genscan was applied. The Genscan-predicted coding sequences may have been edited prior to assembly. (See Example IV.) Alternatively, the identification numbers in column 5 may refer to assemblages of both

10   cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. (See Example V.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon-stretching" algorithm. For example, FL5378618_g4416547_g4091819 is the identification number of a "stretched" sequence, with 5378618 being the Incyte project identification number, g4416547 being the GenBank identification number of

15   the human genomic sequence to which the "exon-stretching" algorithm was applied, and g4091819 being the GenBank identification number of the nearest GenBank protein homolog. (See Example V.) In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

20   Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

25   The invention also encompasses XMES variants. A preferred XMES variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the XMES amino acid sequence, and which contains at least one functional or structural characteristic of XMES.

            The invention also encompasses polynucleotides which encode XMES. In a particular

30   embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:10-18, which encodes XMES. The polynucleotide sequences of SEQ ID NO:10-18, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding XMES. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding XMES. A particular aspect of the invention encompasses a variant of a polynucleotide

5   sequence comprising a sequence selected from the group consisting of SEQ ID NO:10-18 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:10-18. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of XMES.

10   It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding XMES, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in

15   accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring XMES, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode XMES and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring XMES under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding XMES or

20   its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding XMES and its derivatives without altering the encoded amino acid sequences include

25   the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode XMES and XMES derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems

30   using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding XMES or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:10-18 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and

S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

      Methods for DNA sequencing are well known in the art and may be used to practice any of the
5   embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is
10   automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a
15   variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

      The nucleic acid sequences encoding XMES may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences,
20   such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a
25   known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown
30   sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using

commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

5   When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

10   Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate

15   software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which

20   encode XMES may be cloned in recombinant DNA molecules that direct expression of XMES, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express XMES.

The nucleotide sequences of the present invention can be engineered using methods generally

25   known in the art in order to alter XMES-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites,

30   alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or

improve the biological properties of XMES, such as its biological or enzymatic activity or its ability

to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene

variants is produced using PCR-mediated recombination of gene fragments. The library is then

subjected to selection or screening procedures that identify those gene variants with the desired

5   properties. These preferred variants may then be pooled and further subjected to recursive rounds of

DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial"

breeding and rapid molecular evolution. For example, fragments of a single gene containing random

point mutations may be recombined, screened, and then reshuffled until the desired properties are

optimized. Alternatively, fragments of a given gene may be recombined with fragments of

10   homologous genes in the same gene family, either from the same or different species, thereby

maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable

manner.

        In another embodiment, sequences encoding XMES may be synthesized, in whole or in part,

using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids

15   Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively,

XMES itself or a fragment thereof may be synthesized using chemical methods. For example, peptide

synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g.,

Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp.

55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved

20   using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence

of XMES, or any part thereof, may be altered during direct synthesis and/or combined with sequences

from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a

sequence of a naturally occurring polypeptide.

        The peptide may be substantially purified by preparative high performance liquid

25   chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.)

The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing.

(See, e.g., Creighton, supra, pp. 28-53.)

        In order to express a biologically active XMES, the nucleotide sequences encoding XMES or

derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains

30   the necessary elements for transcriptional and translational control of the inserted coding sequence in a

suitable host. These elements include regulatory sequences, such as enhancers, constitutive and

inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences

encoding XMES. Such elements may vary in their strength and specificity. Specific initiation signals

may also be used to achieve more efficient translation of sequences encoding XMES. Such signals

include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding XMES and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous

5 translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

10   Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding XMES and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995)

15 Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

        A variety of expression vector/host systems may be utilized to contain and express sequences encoding XMES. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus);

20 plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO

25 J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di

30 Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

        In bacterial systems, a number of cloning and expression vectors may be selected depending

upon the use intended for polynucleotide sequences encoding XMES. For example, routine cloning,

subcloning, and propagation of polynucleotide sequences encoding XMES can be achieved using a

multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid

(Life Technologies). Ligation of sequences encoding XMES into the vector's multiple cloning site

5   disrupts the lacZ gene, allowing a colorimetric screening procedure for identification of transformed

bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro

transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested

deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem.

264:5503-5509.) When large quantities of XMES are needed, e.g. for the production of antibodies,

10   vectors which direct high level expression of XMES may be used. For example, vectors containing the

strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of XMES. A number of vectors

containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH

promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such

15   vectors direct either the secretion or intracellular retention of expressed proteins and enable integration

of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra;

Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994)

Bio/Technology 12:181-184.)

Plant systems may also be used for expression of XMES. Transcription of sequences encoding

20   XMES may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in

combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311).

Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be

used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science

224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can

25   be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See,

e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp.

191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases

where an adenovirus is used as an expression vector, sequences encoding XMES may be ligated into an

30   adenovirus transcription/translation complex consisting of the late promoter and tripartite leader

sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain

infective virus which expresses XMES in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc.

Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma

virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-

34

based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of XMES in cell lines is preferred. For example, sequences encoding XMES can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk⁻* and *apr⁻* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding XMES is inserted within a marker gene sequence, transformed cells containing sequences encoding XMES can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding XMES under the control of a single

promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding XMES and that express XMES may be identified by a variety of procedures known to those of skill in the art. These

5   procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of XMES using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include

10   enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on XMES is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E.

15   et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding XMES include

20   oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding XMES, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of

25   commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding XMES may be cultured under

30   conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode XMES may be designed to contain signal sequences which direct secretion of XMES through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the

5   protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

10   In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding XMES may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric XMES protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of XMES activity. Heterologous protein and

15   peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, c-myc, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins,

20   respectively. FLAG, c-myc, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the XMES encoding sequence and the heterologous protein sequence, so that XMES may be cleaved away from the heterologous moiety following purification. Methods for fusion protein

25   expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

            In a further embodiment of the invention, synthesis of radiolabeled XMES may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or

30   SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

            XMES of the present invention or fragments thereof may be used to screen for compounds that specifically bind to XMES. At least one and up to a plurality of test compounds may be screened for specific binding to XMES. Examples of test compounds include antibodies, oligonucleotides,

37

proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of XMES, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2):

5   Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which XMES binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express XMES, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E.

10   coli. Cells expressing XMES or cell membrane fractions which contain XMES are then contacted with a test compound and binding, stimulation, or inhibition of activity of either XMES or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example,

15   the assay may comprise the steps of combining at least one test compound with XMES, either in solution or affixed to a solid support, and detecting the binding of XMES to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a

20   solid support.

XMES of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of XMES. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for XMES activity, wherein XMES is combined with at least one test compound, and the activity of XMES in the

25   presence of a test compound is compared with the activity of XMES in the absence of the test compound. A change in the activity of XMES in the presence of the test compound is indicative of a compound that modulates the activity of XMES. Alternatively, a test compound is combined with an in vitro or cell-free system comprising XMES under conditions suitable for XMES activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of XMES

30   may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding XMES or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of

38

human disease. (See, e.g., U.S. Patent Number 5,175,383 and U.S. Patent Number 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989)

5 Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from

10 the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding XMES may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell

15 lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding XMES can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region

20 of a polynucleotide encoding XMES is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress XMES, e.g., by secreting XMES in its milk, may also serve as a

25 convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of XMES and extracellular messengers. In addition, the expression of XMES is closely associated with cartilage and neurological tissues and with tumor tissues from brain, breast,

30 liver, and prostate. Therefore, XMES appears to play a role in neurological disorders, autoimmune/inflammatory disorders, developmental disorders, endocrine disorders, and cell proliferative disorders including cancer. In the treatment of disorders associated with increased XMES expression or activity, it is desirable to decrease the expression or activity of XMES. In the treatment of disorders associated with decreased XMES expression or activity, it is desirable to increase the

expression or activity of XMES.

Therefore, in one embodiment, XMES or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of XMES. Examples of such disorders include, but are not limited to, a neurological disorder

5   such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial

10   thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral

15   palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive

20   dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune

25   polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or

30   pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a developmental disorder such as

40

renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such

5   as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; an endocrine disorder such as a disorder of the hypothalamus and/or pituitary resulting from lesions such as a primary brain tumor, adenoma, infarction associated with pregnancy, hypophysectomy, aneurysm, vascular malformation, thrombosis,

10   infection, immunological disorder, and complication due to head trauma; a disorder associated with hypopituitarism including hypogonadism, Sheehan syndrome, diabetes insipidus, Kallman's disease, Hand-Schuller-Christian disease, Letterer-Siwe disease, sarcoidosis, empty sella syndrome, and dwarfism; a disorder associated with hyperpituitarism including acromegaly, giantism, and syndrome of inappropriate antidiuretic hormone (ADH) secretion (SIADH) often caused by benign adenoma; a

15   disorder associated with hypothyroidism including goiter, myxedema, acute thyroiditis associated with bacterial infection, subacute thyroiditis associated with viral infection, autoimmune thyroiditis (Hashimoto's disease), and cretinism; a disorder associated with hyperthyroidism including thyrotoxicosis and its various forms, Grave's disease, pretibial myxedema, toxic multinodular goiter, thyroid carcinoma, and Plummer's disease; a disorder associated with hyperparathyroidism including

20   Conn disease (chronic hypercalemia); a pancreatic disorder such as Type I or Type II diabetes mellitus and associated complications; a disorder associated with the adrenals such as hyperplasia, carcinoma, or adenoma of the adrenal cortex, hypertension associated with alkalosis, amyloidosis, hypokalemia, Cushing's disease, Liddle's syndrome, and Arnold-Healy-Gordon syndrome, pheochromocytoma tumors, and Addison's disease; a disorder associated with gonadal steroid hormones such as: in

25   women, abnormal prolactin production, infertility, endometriosis, perturbation of the menstrual cycle, polycystic ovarian disease, hyperprolactinemia, isolated gonadotropin deficiency, amenorrhea, galactorrhea, hermaphroditism, hirsutism and virilization, breast cancer, and, in post-menopausal women, osteoporosis; and, in men, Leydig cell deficiency, male climacteric phase, and germinal cell aplasia, a hypergonadal disorder associated with Leydig cell tumors, androgen resistance associated

30   with absence of androgen receptors, 5 α-reductase defficiency, 21-hydroxylase defficiency, and gynecomastia, and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma,

41

teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

In another embodiment, a vector capable of expressing XMES or a fragment or derivative
5 thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of XMES including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified XMES in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of XMES including, but not limited to, those
10 provided above.

In still another embodiment, an agonist which modulates the activity of XMES may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of XMES including, but not limited to, those listed above.

In a further embodiment, an antagonist of XMES may be administered to a subject to treat or
15 prevent a disorder associated with increased expression or activity of XMES. Examples of such disorders include, but are not limited to, those neurological disorders, autoimmune/inflammatory disorders, developmental disorders, endocrine disorders, and cell proliferative disorders including cancer described above. In one aspect, an antibody which specifically binds XMES may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to
20 cells or tissues which express XMES.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding XMES may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of XMES including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary
25 sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with
30 lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of XMES may be produced using methods which are generally known in the art. In particular, purified XMES may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind XMES. Antibodies to XMES may also be generated using methods that are well known in the art. Such antibodies may include, but are not

limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, 5   and others may be immunized by injection with XMES or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG 10   (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to XMES have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of XMES amino 15   acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to XMES may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma 20   technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate 25   antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce XMES-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be 30   generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter,

G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for XMES may also be generated. For example, such fragments include, but are not limited to, F(ab')$_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2

5   fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either

10   polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between XMES and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering XMES epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

15   Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for XMES. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of XMES-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple

20   XMES epitopes, represents the average affinity, or avidity, of the antibodies for XMES. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular XMES epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the XMES-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from

25   about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of XMES, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine

30   the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of XMES-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and

Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding XMES, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA,

5   PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding XMES. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding XMES. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense

10   sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) J. Allergy Cli. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral

15   vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res.

20   25(14):2730-2736.)   ·

In another embodiment of the invention, polynucleotides encoding XMES may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined

25   immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal,

30   R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399),

hepatitis B or C virus (HBV, HCV); fungal parasites, such as <u>Candida albicans</u> and <u>Paracoccidioides</u> <u>brasiliensis</u>; and protozoan parasites such as <u>Plasmodium falciparum</u> and <u>Trypanosoma cruzi</u>). In the case where a genetic deficiency in XMES expression or regulation causes disease, the expression of XMES from an appropriate population of transduced cells may alleviate the clinical manifestations

5   caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in XMES are treated by constructing mammalian expression vectors encoding XMES and introducing these vectors by mechanical means into XMES-deficient cells. Mechanical transfer technologies for use with cells <u>in vivo</u> or <u>ex vitro</u> include (i) direct DNA microinjection into individual cells, (ii) ballistic gold

10   particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of XMES include, but are not

15   limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). XMES may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the

20   tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V.

25   and Blau, H.M. <u>supra</u>)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding XMES from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental

30   parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with

46

respect to XMES expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding XMES under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences

5   required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al.

10   (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference.

15   Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4$^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

20   In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding XMES to cells which have one or more genetic abnormalities with respect to the expression of XMES. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas

25   (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

30   In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding XMES to target cells which have one or more genetic abnormalities with respect to the expression of XMES. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing XMES to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with

ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has

been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res.

169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S.

Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is

5   hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV

d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under

the control of the appropriate promoter for purposes including human gene therapy. Also taught by this

patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22.

For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev.

10   Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus

sequences, the generation of recombinant virus following the transfection of multiple plasmids

containing different segments of the large herpesvirus genomes, the growth and propagation of

herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary

skill in the art.

15   In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to

deliver polynucleotides encoding XMES to target cells. The biology of the prototypic alphavirus,

Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on

the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During

alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid

20   proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting

in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g.,

protease and polymerase). Similarly, inserting the coding sequence for XMES into the alphavirus

genome in place of the capsid-coding region results in the production of a large number of XMES-

coding RNAs and the synthesis of high levels of XMES in vector transduced cells. While alphavirus

25   infection is typically associated with cell lysis within a few days, the ability to establish a persistent

infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that

the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application

(Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the

introduction of XMES into a variety of cell types. The specific transduction of a subset of cells in a

30   population may require the sorting of cells prior to transduction. The methods of manipulating

infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and

performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10

and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can

48

be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr,

5   Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme

10   molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding XMES.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA,

15   GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

20   Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding XMES. Such DNA sequences may be incorporated into a wide variety of vectors with

25   suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends

30   of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding XMES. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming

5   oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased XMES expression or activity, a compound which specifically inhibits expression of the

10   polynucleotide encoding XMES may be therapeutically useful, and in the treatment of disorders associated with decreased XMES expression or activity, a compound which specifically promotes expression of the polynucleotide encoding XMES may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method

15   commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a

20   polynucleotide encoding XMES is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an _in vitro_ cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding XMES are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence

25   of the polynucleotide encoding XMES. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific

30   polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide

35   nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide

sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors may be introduced into stem cells taken

5   from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such

10   therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various

15   formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of XMES, antibodies to XMES, and mimetics, agonists, antagonists, or inhibitors of XMES.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal,

20   intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting

25   formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

30   Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising XMES or fragments thereof. For example, liposome preparations

containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, XMES or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et

5   al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for

10   administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example XMES or fragments thereof, antibodies of XMES, and agonists, antagonists or inhibitors of XMES, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by

15   calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is

20   preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active

25   moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

30   Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells,

conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind XMES may be used for the diagnosis of disorders characterized by expression of XMES, or in assays to monitor patients being

5   treated with XMES or agonists, antagonists, or inhibitors of XMES. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for XMES include methods which utilize the antibody and a label to detect XMES in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter

10   molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring XMES, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of XMES expression. Normal or standard values for XMES expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to XMES under

15   conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of XMES expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding XMES may be used for

20   diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of XMES may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of XMES, and to monitor regulation of XMES levels during therapeutic intervention.

25   In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding XMES or closely related molecules may be used to identify nucleic acid sequences which encode XMES. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the

30   probe identifies only naturally occurring sequences encoding XMES, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the XMES encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:10-18 or from

53

genomic sequences including promoters, enhancers, and introns of the XMES gene.

Means for producing specific hybridization probes for DNAs encoding XMES include the cloning of polynucleotide sequences encoding XMES or XMES derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may

5   be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding XMES may be used for the diagnosis of disorders

10   associated with expression of XMES. Examples of such disorders include, but are not limited to, a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other

15   demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal

20   syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety,

25   and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis,

30   anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis,

hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or
pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's
syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus
erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner

5 syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial,
fungal, parasitic, protozoal, and helminthic infections, and trauma; a developmental disorder such as
renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker
muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia,
genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic

10 syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such
as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure
disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis,
congenital glaucoma, cataract, and sensorineural hearing loss; an endocrine disorder such as a disorder
of the hypothalamus and/or pituitary resulting from lesions such as a primary brain tumor, adenoma,

15 infarction associated with pregnancy, hypophysectomy, aneurysm, vascular malformation, thrombosis,
infection, immunological disorder, and complication due to head trauma; a disorder associated with
hypopituitarism including hypogonadism, Sheehan syndrome, diabetes insipidus, Kallman's disease,
Hand-Schuller-Christian disease, Letterer-Siwe disease, sarcoidosis, empty sella syndrome, and
dwarfism; a disorder associated with hyperpituitarism including acromegaly, giantism, and syndrome of

20 inappropriate antidiuretic hormone (ADH) secretion (SIADH) often caused by benign adenoma; a
disorder associated with hypothyroidism including goiter, myxedema, acute thyroiditis associated with
bacterial infection, subacute thyroiditis associated with viral infection, autoimmune thyroiditis
(Hashimoto's disease), and cretinism; a disorder associated with hyperthyroidism including
thyrotoxicosis and its various forms, Grave's disease, pretibial myxedema, toxic multinodular goiter,

25 thyroid carcinoma, and Plummer's disease; a disorder associated with hyperparathyroidism including
Conn disease (chronic hypercalemia); a pancreatic disorder such as Type I or Type II diabetes mellitus
and associated complications; a disorder associated with the adrenals such as hyperplasia, carcinoma,
or adenoma of the adrenal cortex, hypertension associated with alkalosis, amyloidosis, hypokalemia,
Cushing's disease, Liddle's syndrome, and Arnold-Healy-Gordon syndrome, pheochromocytoma

30 tumors, and Addison's disease; a disorder associated with gonadal steroid hormones such as: in
women, abnormal prolactin production, infertility, endometriosis, perturbation of the menstrual cycle,
polycystic ovarian disease, hyperprolactinemia, isolated gonadotropin deficiency, amenorrhea,
galactorrhea, hermaphroditism, hirsutism and virilization, breast cancer, and, in post-menopausal
women, osteoporosis; and, in men, Leydig cell deficiency, male climacteric phase, and germinal cell

aplasia, a hypergonadal disorder associated with Leydig cell tumors, androgen resistance associated with absence of androgen receptors, 5 α-reductase defficiency, 21-hydroxylase deficiency, and gynecomastia, and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis,

5   paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

10   The polynucleotide sequences encoding XMES may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered XMES expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding XMES may be useful in assays that

15   detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding XMES may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control

20   sample then the presence of altered levels of nucleotide sequences encoding XMES in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of XMES,

25   a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding XMES, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used.

30   Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the

56

patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding XMES may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding XMES, or a fragment of a polynucleotide complementary to the polynucleotide encoding XMES, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding XMES may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding XMES are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of XMES include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be

5   accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray

10   can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the

15   activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, XMES, fragments of XMES, or antibodies specific for XMES may be

20   used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by

25   quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the

30   hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression _in vivo_, as in the

58

case of a tissue or biopsy sample, or <u>in vitro</u>, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with <u>in vitro</u> model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and

analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are

5   visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein

10   spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

15   A proteomic profile may also be generated using antibodies specific for XMES to quantify the levels of XMES expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of

20   methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

             Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J.

25   Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

             In another embodiment, the toxicity of a test compound is assessed by treating a biological

30   sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the

individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized

5 by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci.

10 USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

15 In another embodiment of the invention, nucleic acid sequences encoding XMES may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal

20 mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet.

25 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map

30 data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding XMES on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is

5   valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may

10   also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, XMES, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a

15   solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between XMES and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are

20   synthesized on a solid substrate. The test compounds are reacted with XMES, or fragments thereof, and washed. Bound XMES is then detected by methods well known in the art. Purified XMES can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

25   In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding XMES specifically compete with a test compound for binding XMES. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with XMES.

In additional embodiments, the nucleotide sequences which encode XMES may be used in any

30   molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are,

62

therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, including U.S. Ser. No. 60/210,233, U.S. Ser. No. 60/213,465, and U.S. Ser. No. 60/249,019 are

5   expressly incorporated by reference herein.

EXAMPLES

I.   Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database
10   (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and
15   ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was
20   isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the
25   recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column
30   chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), or pINCY (Incyte Genomics, Palo Alto CA), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells

including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX
DH10B from Life Technologies.

II.   Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision

5   using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least
one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC
Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL
8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid
purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of

10   distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a
high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal
cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-
well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using

15   PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner
(Labsystems Oy, Helsinki, Finland).

III.   Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows.
Sequencing reactions were processed using standard methods or high-throughput instrumentation

20   such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler
(MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the
MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared
using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as
the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

25   Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were
carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI
PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI
protocols and base calling software; or other sequence analysis systems known in the art. Reading
frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997,

30   supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques
disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector,
linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based
on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA

64

sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families.

5   See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length.

10   Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently

15   analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are

20   generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

        Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold

25   parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity

30   between two sequences).

        The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:10-18. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

IV.   Identification and Editing of Coding Sequences from Genomic DNA

Putative extracellular messengers were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode extracellular messengers, the encoded polypeptides were analyzed by querying against PFAM models for extracellular messengers. Potential extracellular messengers were also identified by homology to Incyte cDNA sequences that had been annotated as extracellular messengers. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V.   Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to

66

be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were

5   given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

10   <u>"Stretched" Sequences</u>

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis

15   to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the

20   public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

            VI.   **Chromosomal Mapping of XMES Encoding Polynucleotides**

The sequences which were used to assemble SEQ ID NO:10-18 were compared with

25   sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:10-18 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for

30   Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-

arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation

5   hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

In this manner, SEQ ID NO:16 was mapped to chromosome 2 within the interval from

10   190.80 to 197.60 centiMorgans.

VII.   Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995)

15   supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar.

20   The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score x Percent Identity}}{5 \text{ x minimum } \{\text{length(Seq. 1), length(Seq. 2)}\}}$$

25   The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and –4 for

30   every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and

70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding XMES are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least

5 in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory

10 system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all

15 categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding XMES. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of XMES Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate

20 fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence

25 at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR

30 was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C,

2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

5   The concentration of DNA in each well was determined by dispensing 100 μl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 μl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the

10   concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For

15   shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing

20   media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4

25   repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle

30   sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX.   Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:10-18 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National

5   Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human

10 genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature

15 under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

X.   Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing

20 photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra.), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a

25 substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

30 Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection.

After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on
5   the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is
10   reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/µl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/µl RNase inhibitor, 500 µM dATP, 500 µM dGTP, 500 µM dTTP, 40 µM dCTP, 40 µM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by _in vitro_ transcription
15   from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated
20   using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 µl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is
25   amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 µg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

30   Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C
35   oven.

72

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1 µl of the array element DNA, at an average concentration of 100 ng/µl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

5   Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

10   **Hybridization**

Hybridization reactions contain 9 µl of sample mixture consisting of 0.2 µg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly
15   larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 µl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

Detection

20   Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-
25   scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate
30   filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a
35   cDNA control species added to the sample mixture at a known concentration. A specific location on

the array contains a complementary DNA sequence, allowing the intensity of the signal at that

location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples

from different sources (e.g., representing test and control cells), each labeled with a different

fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially

5   expressed, the calibration is done by labeling samples of the calibrating cDNA with the two

fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital

(A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC

computer. The digitized data are displayed as an image where the signal intensity is mapped using a

10   linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high

signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and

measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping

emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot

15   is centered in each element of the grid. The fluorescence signal within each element is then integrated

to obtain a numerical value corresponding to the average intensity of the signal. The software used

for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

XI.   Complementary Polynucleotides

Sequences complementary to the XMES-encoding sequences, or any parts thereof, are used to

20   detect, decrease, or inhibit expression of naturally occurring XMES. Although use of oligonucleotides

comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with

smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO

4.06 software (National Biosciences) and the coding sequence of XMES. To inhibit transcription, a

complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent

25   promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is

designed to prevent ribosomal binding to the XMES-encoding transcript.

XII.   Expression of XMES

Expression and purification of XMES is achieved using bacterial or virus-based expression

systems. For expression of XMES in bacteria, cDNA is subcloned into an appropriate vector

30   containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA

transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid

promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory

element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3).

Antibiotic resistant bacteria express XMES upon induction with isopropyl beta-D-

thiogalactopyranoside (IPTG). Expression of XMES in eukaryotic cells is achieved by infecting insect

or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus

(AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is

replaced with cDNA encoding XMES by either homologous recombination or bacterial-mediated

5   transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong

polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to

infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases.

Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et

al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther.

10   7:1937-1945.)

            In most expression systems, XMES is synthesized as a fusion protein with, e.g., glutathione S-

transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step,

affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton

enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized

15   glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia

Biotech). Following purification, the GST moiety can be proteolytically cleaved from XMES at

specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification

using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-

His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins

20   (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra,

ch. 10 and 16). Purified XMES obtained by these methods can be used directly in the assays shown in

Examples XVI, and XVII where applicable.

XIII.   Functional Assays

            XMES function is assessed by expressing the sequences encoding XMES at physiologically

25   elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression

vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice

include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which

contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a

human cell line, for example, an endothelial or hematopoietic cell line, using either liposome

30   formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a

marker protein are co-transfected. Expression of a marker protein provides a means to distinguish

transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the

recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP;

Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-

based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in

5 cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G.

10 (1994) Flow Cytometry, Oxford, New York NY.

The influence of XMES on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding XMES and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using

15 magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding XMES and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIV.   Production of XMES Specific Antibodies

20   XMES substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the XMES amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is

25 synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-

30 Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-XMES activity by, for example, binding the peptide or XMES to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

76

XV.   Purification of Naturally Occurring XMES Using Specific Antibodies

Naturally occurring or recombinant XMES is substantially purified by immunoaffinity chromatography using antibodies specific for XMES. An immunoaffinity column is constructed by covalently coupling anti-XMES antibody to an activated chromatographic resin, such as
5 CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing XMES are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of XMES (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt
10 antibody/XMES binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and XMES is collected.

XVI.   Identification of Molecules Which Interact with XMES

XMES, or biologically active fragments thereof, are labeled with [125]I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules
15 previously arrayed in the wells of a multi-well plate are incubated with the labeled XMES, washed, and any wells with labeled XMES complex are assayed. Data obtained using different concentrations of XMES are used to calculate values for the number, affinity, and association of XMES with the candidate molecules.

Alternatively, molecules interacting with XMES are analyzed using the yeast two-hybrid
20 system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

XMES may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent
25 No. 6,057,101).

XVII.   Demonstration of XMES Activity

An assay for XMES-1 activity measures its inhibitory activity on Hepatocyte Growth Factor (HGF) activator. In this assay, HGF activator (450 ng/ml) is mixed with various concentrations of purified XMES-1 in PBS containing 0.05% CHAPS and incubated at 37 degrees C for 30 minutes to
30 form an enzyme-inhibitor complex. The remaining HGF-converting activity in the mixture is measured by the addition of equal amounts of single chain HGF (sc-HGF) (1.5 μg/ml in PBS containing 0.05% CHAPS) and dextran sulfate (100 mg/ml, MWCO=500,000, Sigma) followed by further incubation for 2 hours, and subsequent analysis by SDS-PAGE under reducing gel conditions. The gel is stained with coomassie blue and the amounts of sc-HGF and the heterodimeric form are measured by scanning the

stained bands. The inhibitory activity of XMES-1 against HGF activator is estimated by calculating the ratio of the remaining single chain form to total HGF (Shimomura, T. et al. (1997) J. Biol. Chem. 272:6370-6376).

An assay for XMES activity measures cell proliferation as the amount of newly initiated DNA synthesis in Swiss mouse 3T3 cells. A plasmid containing polynucleotides encoding XMES is transfected into quiescent 3T3 cultured cells using methods well known in the art. The transiently transfected cells are then incubated in the presence of [^{3}H]thymidine, a radioactive DNA precursor. Where applicable, varying amounts of XMES ligand are added to the transfected cells. Incorporation of [^{3}H]thymidine into acid-precipitable DNA is measured over an appropriate time interval, and the amount incorporated is directly proportional to the amount of newly synthesized DNA.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
1657368	1	1657368CD1	10	1657368CB1
4028972	2	4028972CD1	11	4028972CB1
5398353	3	5398353CD1	12	5398353CB1
71234118	4	71234118CD1	13	71234118CB1
240168	5	240168CD1	14	240168CB1
7481107	6	7481107CD1	15	7481107CB1
7476245	7	7476245CD1	16	7476245CB1
5819744	8	5819744CD1	17	5819744CB1
5378618	9	5378618CD1	18	5378618CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
1	1657368CD1	g3721617	3.30E-42	[Mus musculus] mac25 insulin-like growth factor binding protein Komatsu S. et al. (2000) Biochem Biophys Res Commun 267:109-117
		g8217418	1.00E-169	[Homo sapiens] bA108L7.1 (novel Insulin-like growth factor binding type protein with Kazal-type serine protease inhibitor domain)
2	4028972CD1	g294600	0	[Rattus norvegicus] neurexin III-alpha Ushkaryov Y.A. and Sudhof T.C. (1993) Proc Natl Acad Sci U S A 1993 90:6410-6414
3	5398353CD1	g4514718	1.50E-22	FGF-19 [Homo sapiens] (Nishimura, T. (1999) Biochim. Biophys. Acta 1444:148-151)
		g9049445	1.00E-120	[Homo sapiens] FGF-21
4	71234118CD1	g11177164	4.00E-59	[Mus musculus] polydom protein
		g1438954	3.20E-18	[Homo sapiens] neuronal pentraxin 1 Omeis, I.A. et al. (1996) Genomics 36:543-545
5	240168CD1	g2924601	1.10E-29	[Homo sapiens] hepatocyte growth factor activator inhibitor Shimomura, T. et al. (1997) J. Biol. Chem. 272:6370-6376

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
6	7481107CD1	g632490	7.50E-185	[Bos taurus] cartilage-derived morphogenetic protein Chang, S.C. et al. (1994) J. Biol. Chem. 269:28227-28234
7	7476245CD1	g189228	2.70E-54	[Homo sapiens] neuromedin B Krane, I.M. et al. (1988) J. Biol. Chem. 263:13317-13323
9	5378618CD1	g4590406	2.90E-19	[Drosophila melanogaster] slit protein Kidd, T. et al. (1999) Cell 96:785-794
		g9309467	0	[Mus musculus] leucine-rich glioma-inactivated 1 protein precursor

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	1657368CD1	304	T250 Y251 T112 S117 T185	N159 N183 N277	M1-A30 Signal peptide	HMMER
					Signal_cleavage: M1-A30	SPSCAN
					C53-C90 Insulin-like growth factor binding proteins	HMMER_PFAM
					V121-C168 Kazal-type serine protease inhibitor domain	HMMER_PFAM
					G186-G255 Immunoglobulin domain	HMMER_PFAM
					G50-D115 Insulin-like growth factor binding proteins signature	PROFILESCAN
					L85-G171 Kazal serine protease inhibitors family signature	PROFILESCAN
					V184-L271 MAC25 FOLLISTATINLIKE PROSTACYCLIN STIMULATING FACTOR	BLAST_PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
2	4028972CD1	1438	T11 S46 S100 T264 S290 T303 T469 S503 T531 T537 S564 S647 S714 S719 T730 T737 S762 T782 S791 T823 T878 S949 S972 T47 S79 S282 S283 T372 S392 T507 T708 T730 T795 T810 T906 Y273 Y547 Y770 Y860	N58 N105 N761 N403 N592 N776 N940 N943	C662-C673 Asx_Hydroxyl	MOTIFS
					S20-M31 Gatase_Type_I	MOTIFS
					signal_cleavage: M1-G27	SPSCAN
					F10-G27, M1365-A1383 Transmembrane domain	HMMER
					F55-G174, F291-R420, F479-N626, L740-I850, F903-D1034, F1126-W1201 Laminin G domain	HMMER_PFAM
					C202-C234, C651-C683, C1057-C1089 EGF-like domain	HMMER_PFAM
					S647-N658 Type II EGF-like signature	BLIMPS_PRINTS
					E204-C224 COMPLEMENT C9 SIGNATURE	BLIMPS_PRINTS
					F1098-Y1437 II-BETA NEUREXIN	BLAST_DOMO
					F717-E802, G803-Q902, G1210-T1338, T1338-V1438 ALTERNATIVE NEUREXIN PRECURSOR	BLAST_PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
3	5398353CD1	208	S121 T6 T97		signal_cleavage: M1-A27	SPSCAN
					Fibroblast growth factor FGF:H59-H139	HMMER_PFAM
					HEPARIN BINDING GROWTH FACTOR FAMILY PR00263B:L80-G94 PR00263C:F100-S112 PR00263D:P117-S136	BLIMPS_PRINTS
					IL1/HBGF FAMILY SIGNATURE PR00262A:P87-H114 PR00262B:A119-H139	BLIMPS_PRINTS
					HBGF/FGF family signature hbgf_fgf.prf:P87-G140	PROFILESCAN
					GROWTH FACTOR FIBROBLAST MITOGEN SIGNAL HEPARIN BINDING VASCULARIZATION PD000831:H59-H139	BLAST_PRODOM
					HBGF/FGF FAMILY DM00642│P48803│47-205:S16- P165 DM00642│P48805│29-186:D32-P165 DM00642│P10767│50-207:G41-H139 DM00642│P05524│5-181:G40-P160	BLAST_DOMO

84

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
4	7123418CD1	159	S114 S2 S69 Y75	N6 N85	signal_cleavage: M1-G35	SPSCAN
					Pentaxin family pentaxin: S80-K141	HMMER_PFAM
					Pentaxin family proteins BL00289C:H122-G140	BLIMPS_BLOCKS
					Pentaxin family signature pentaxin.prf:V100-A147	PROFILESCAN
					PENTAXIN SIGNATURE PR00895A:L58-D72 PR00895C:H122-G140	BLIMPS_PRINTS
					PRECURSOR SIGNAL PENTAXIN GLYCOPROTEIN PLASMA CREACTIVE CALCIUM ACUTE PHASE PD002153:Y48-G157	BLAST_PRODOM
					C-REACTIVE PROTEIN DM00835|P47971|194-431:F36-I155 DM00835|P47970|187-426:N37-I155	BLAST_DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
5	240168CD1	500	S11 S201 S234 S339 S34 S377 S382 S395 S422 S435 S53 T166 T197 T263 T279 T397 T487 Y431	N164 N291 N401	LDL RECEPTOR LIGAND-BINDING REPEAT DM00045|P98072|663-717:H314-L347	BLAST_DOMO
					HEPATOCYTE GROWTH FACTOR ACTIVATOR INHIBITOR GLYCOPROTEIN PD120361:G84-L297, S11-P44	BLAST_PRODOM
					LDL-receptor class A (LDLRA) domain proteins BL01209:C329-E341	BLIMPS_BLOCKS
					LOW DENSITY LIPOPROTEIN PR00261:G320-E341	BLIMPS_PRINTS
					signal peptide signal_peptide:M1-A37 transmembrane domain transmem_domain:V452-C471	HMMER
					Low-density lipoprotein receptor domain ldl_recept_a:H308-N346	HMMER_PFAM
					Spscan signal_cleavage:M1-A37	SPSCAN

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
6	7481107CD1	455	S106 S120 S129 S156 S284 S33 S355 S39 S393 S45 T3 T336	N114	TGF-BETA FAMILY DM00245|P43026|174-501:R331-R455, L99-G298 DM00245|P12644|95-408:R332-R455, A124-V217 DM00245|P12643|88-396:R345-R455, Q119-V217 DM00245|I49541|105-420:R333-R455, A124-V217	BLAST_DOMO
					GLYCOPROTEIN PRECURSOR SIGNAL GROWTH FACTOR PROTEIN CYTOKINE BETA BONE MORPHOGENETIC PD000357:C354-R455	BLAST_PRODOM
					TGF-beta family proteins BL00250:C354-F389, C419-C454	BLIMPS_BLOCKS
					GROWTH FACTOR CYSTINE KNOT SUPERFAMILY SIGNATURE PR00438:E379-D388, E450-C454	BLIMPS_PRINTS
					INHIBIN ALPHA CHAIN SIGNATURE PR00669:C354-W371	
					signal peptide signal peptide:M1-G22	HMMER
					Transforming growth factor beta like TGF-beta:C354-R455	HMMER_PFAM
					TGF-beta propeptide TGFb propeptide:G61-R268	
					Tgf_beta:I372-C387	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites.	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
7	7476245CD1	121	S76		BOMBESIN-LIKE PEPTIDES FAMILY DM00828|P08949|18-55:A18-M56	BLAST_DOMO
					DM00828|P01297|1-31:A25-M56 NEUROMEDIN B32 PRECURSOR	BLAST_PRODOM
					CONTAINS: B BOMBESIN FAMILY AMIDATION CLEAVAGE ON PAIR OF BASIC RESIDUES SIGNAL PD054439:G57-K121	
					NEUROMEDIN B32 CONTAINS: B BOMBESIN FAMILY AMIDATION PRECURSOR CLEAVAGE ON PD026110:A25-M56	
					Bombesin-like peptides family proteins BL00257:R46-M56	BLIMPS_BLOCKS
					signal peptide signal_peptide:M1-D30	HMMER
					Bombesin:W50-M56	MOTIFS
					Bombesin-like peptides family signature bombesin.prf:D30-R81	PROFILESCAN
					Spscan signal_cleavage:M1-P26	SPSCAN
8	5819744CD1	55	S12 S17 T2		Parathyroid hormone family signature parathyroid.prf:M1-K50	PROFILESCAN
					Spscan signal_cleavage:M1-P51	SPSCAN

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9	5378618CD1	545	S154 S158 S238 S345 S366 S390 S404 S405 S425 S471 S528 T188 T209 T216 T374 T518 T72	N186 N271 N402 N70	Leucine-rich repeat signature PR00019:L156-F169, L135-L148	BLIMPS_PRINTS
					signal peptide signal_peptide:M1-A28	HMMER
					Leucine Rich Repeat LR:D62-P85, S86-F109, H110-R133, D134-D157	HMMER_PFAM
					Leucine rich repeat C-terminal domain LRRCT:N167-T216	
					Spscan signal_cleavage:M1-L23	SPSCAN
					Atp_Gtp_A binding site:A526-T533	MOTIFS

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
10	1657368CB1	1374	1279-1374	7185657H1 (BONRFEC01)	32	571
11	4028972CB1	4541	916-2416, 1-158, 2747-3709	1657373H1 (URETTUT01)	1	227
				2687816F6 (LUNGNOT23)	833	1374
				5851853H1 (FIBAUNT02)	684	942
				GNN.g6967325_000005_006	199	1113
				2304926R6 (NGANNOT01)	4158	4541
				70032782D1	3771	4163
				70032296D1	3418	3776
				5498652F6 (BRABDIR01)	3659	4154
				70503374V1	516	1173
				70503718V1	2106	2675
				6608073H1 (BRAXDIC01)	3063	3519
				70505915V1	2705	3333
				70505907V1	1413	2072
				6993321H1 (BRAQTDR02)	1	561
				70501232V1	784	1409
				70504786V1	1385	2066
				70504213V1	1997	2655
				70505588V1	2429	3132
12	5398353CB1	1117	485-916, 1-75, 990-1117, 157-211	71281268V1	1	631
				71152050V1	682	1117
				71154676V1	657	1114
				5398353T6 (LIVRTUT13)	506	1084
13	71234118CB1	2460	2070-2175, 1-38, 1692-1726, 1395-1534, 156-1013, 2409-2460	4062433T6 (BRAINOT21)	1798	2460
				71008692V1	219	836
				71234485V1	1	590
				71010865V1	1150	1746
				5647010F8 (BRAITUT23)	1727	2360
				71009374V1	978	1560
				71008960V1	599	1085

Table 4 (cont.)

Poly-nucleotide SEQ ID NO:	Incyte Poly-nucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
14	240168CB1	2601	592-648, 1-45, 706-754	3227038H1 (EPIGNOT01)	1348	1626
				g3153915	649	1174
				1833192H1 (BRAINON01)	2367	2601
				1750805F6 (LIVRTUT01)	2189	2580
				676413OJ1 (BRAUNOR01)	657	1344
				1913790H1 (PROSTUT04)	2340	2587
				6975777H1 (BRAHTDR04)	1357	2035
				7070876H1 (BRAUTDR02)	1058	1545
				7653484H1 (UTREDME06)	1	471
				GNN.g7768040_000033_002	147	759
				1451206F6 (PENITUT01)	1618	2210
15	7481107CB1	2791	1-1312, 2425-2791, 1841-1883, 1598-1658	GNN.g8439934_000008_002.edit.1	1029	2396
				1387140F6 (CARGDIT02)	2188	2791
				6152346H1 (ENDMUNT04)	2165	2451
				GBI.chr8_smoosh_4.comp	1	1427
16	7476245CB1	709	651-709	2377775F6 (ISLTNOT01)	411	709
				1577577F6 (LNODNOT03)	1	554
17	5819744CB1	753	1-753	5813552T8 (PROSTUS23)	210	753
				5813552F8 (PROSTUS23)	1	636
18	5378618CB1	3413	3392-3413, 1-1039, 1696-2626	6718171H1 (CONDTUT02)	1805	2358
				70656442V1	2901	3413
				GNN.g4416547_000001_4	78	1715
				8264863U1	1683	2349
				FL5378618_g4416547_g4091 819	78	1715
				GBI:g4416547.fasta.edit.1	1	201
				70657537V1	2340	2945

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
10	1657368CB1	LUNGNOT23
11	4028972CB1	BRABDIR01
12	5398353CB1	LIVRTUT13
13	71234118CB1	PLACNOB01
14	240168CB1	LIVRTUT01
15	7481107CB1	CARGDIT02
16	7476245CB1	BRAITUT12
17	5819744CB1	PROSTUS23
18	5378618CB1	BRSTTUT03

Table 6

Library	Vector	Library Description
BRABDIR01	pINCY	Library was constructed using RNA isolated from diseased cerebellum tissue removed from the brain of a 57-year-old Caucasian male, who died from a cerebrovascular accident. Patient history included Huntington's disease, emphysema, and tobacco abuse.
BRAITUT12	pINCY	Library was constructed using RNA isolated from brain tumor tissue removed from the left frontal lobe of a 40-year-old Caucasian female during excision of a cerebral meningeal lesion. Pathology indicated grade 4 gemistocytic astrocytoma.
BRSTTUT03	PSPORT1	Library was constructed using RNA isolated from breast tumor tissue removed from a 58-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated multicentric invasive grade 4 lobular carcinoma. The mass was identified in the upper outer quadrant, and three separate nodules were found in the lower outer quadrant of the left breast. Patient history included skin cancer, rheumatic heart disease, osteoarthritis, and tuberculosis. Family history included cerebrovascular disease, coronary artery aneurysm, breast cancer, prostate cancer, atherosclerotic coronary artery disease, and type I diabetes.
CARGDIT02	pINCY	Library was constructed using RNA isolated from cartilage obtained from 4 donors with end-stage osteoarthritis. The patients had received a variety of non-steroidal anti-inflammatory drugs.
LIVRTUT01	pINCY	Library was constructed using RNA isolated from liver tumor tissue removed from a 51-year-old Caucasian female during a hepatic lobectomy. Pathology indicated metastatic grade 3 adenocarcinoma consistent with colon cancer. Family history included a malignant neoplasm of the liver.

Table 6 (cont.)

Library	Vector	Library Description
LIVRTUT13	pINCY	Library was constructed using RNA isolated from liver tumor tissue removed from a 62-year-old Caucasian female during partial hepatectomy and exploratory laparotomy. Pathology indicated metastatic intermediate grade neuroendocrine carcinoma, consistent with islet cell tumor, forming nodules ranging in size, in the lateral and medial left liver lobe. The pancreas showed fibrosis, chronic inflammation and fat necrosis consistent with pseudocyst. The gall bladder showed mild chronic cholecystitis. Patient history included malignant neoplasm of the pancreas tail, pulmonary embolism, hyperlipidemia, thrombophlebitis, joint pain in multiple joints, type II diabetes, benign hypertension, and cerebrovascular disease. Family history included pancreas cancer, secondary liver cancer, benign hypertension, and hyperlipidemia.
LUNGNOT23	pINCY	Library was constructed using RNA isolated from left lobe lung tissue removed from a 58-year-old Caucasian male. Pathology for the associated tumor tissue indicated metastatic grade 3 (of 4) osteosarcoma. Patient history included soft tissue cancer, secondary cancer of the lung, prostate cancer, and an acute duodenal ulcer with hemorrhage. Family history included prostate cancer, breast cancer, and acute leukemia.
PLACNOB01	PBLUESCRIPT	Library was constructed using RNA isolated from placenta.

Table 6 (cont.)

Library	Vector	Library Description
PROSTUS23	pINCY	This subtracted prostate tumor library was constructed using 10 million clones from a pooled prostate tumor library that was subjected to 2 rounds of subtractive hybridization with 10 million clones from a pooled prostate tissue library. The starting library for subtraction was constructed by pooling equal numbers of clones from 4 prostate tumor libraries using mRNA isolated from prostate tumor removed from Caucasian males at ages 58 (A), 61 (B), 66 (C), and 68 (D) during prostatectomy with lymph node excision. Pathology indicated adenocarcinoma in all donors. History included elevated PSA, induration and tobacco abuse in donor A; elevated PSA, induration, prostate hyperplasia, renal failure, osteoarthritis, renal artery stenosis, benign HTN, thrombocytopenia, hyperlipidemia, tobacco/alcohol abuse and hepatitis C (carrier) in donor B; elevated PSA, induration, and tobacco abuse in donor C; and elevated PSA, induration, hypercholesterolemia, and kidney calculus in donor D. The hybridization probe for subtraction was constructed by pooling equal numbers of cDNA clones from 3 prostate tissue libraries derived from prostate tissue, prostate epithelial cells, and fibroblasts from prostate stroma from 3 different donors. Subtractive hybridization conditions were based on the methodologies of Swaroop et al., NAR 19 (1991):1954 and Bonaldo, et al. Genome Research 6 (1996):791.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	*ESTs*: Probability value= 1.0E-8 or less *Full Length sequences*: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs*: fasta E value=1.06E-6 *Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences*: fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	*PFAM hits*: Probability value= 1.0E-3 or less *Signal peptide hits*: Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score>GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence selected from the group consisting of
5   SEQ ID NO:1-9,

b) a naturally occurring polypeptide comprising an amino acid sequence at least 90%
identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-9,

c) a biologically active fragment of a polypeptide having an amino acid sequence selected
from the group consisting of SEQ ID NO:1-9, and

10   d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from
the group consisting of SEQ ID NO:1-9.

2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-9.

15   3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID
20   NO:10-18.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a
polynucleotide of claim 3.

25   7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method for producing a polypeptide of claim 1, the method comprising:

30   a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said
cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide
comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim
1, and

b) recovering the polypeptide so expressed.

35

10. An isolated antibody which specifically binds to a polypeptide of claim 1.

11. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting

5   of SEQ ID NO:10-18,

b) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90%

identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:10-18,

c) a polynucleotide complementary to a polynucleotide of a),

d) a polynucleotide complementary to a polynucleotide of b), and

10   e) an RNA equivalent of a)-d).

12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a

polynucleotide of claim 11.

15   13. A method for detecting a target polynucleotide in a sample, said target polynucleotide

having a sequence of a polynucleotide of claim 11, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides

comprising a sequence complementary to said target polynucleotide in the sample, and which probe

specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization

20   complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if

present, the amount thereof.

14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.

25

15. A method for detecting a target polynucleotide in a sample, said target polynucleotide

having a sequence of a polynucleotide of claim 11, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction

amplification, and

30   b) detecting the presence or absence of said amplified target polynucleotide or fragment

thereof, and, optionally, if present, the amount thereof.

16. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable

excipient.

35

17. A composition of claim 16, wherein the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO:1-9.

18. A method for treating a disease or condition associated with decreased expression of functional XMES, comprising administering to a patient in need of such treatment the composition of claim 16.

19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

21. A method for treating a disease or condition associated with decreased expression of functional XMES, comprising administering to a patient in need of such treatment a composition of claim 20.

22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

24. A method for treating a disease or condition associated with overexpression of functional XMES, comprising administering to a patient in need of such treatment a composition of claim 23.

25. A method of screening for a compound that specifically binds to the polypeptide of claim 1, said method comprising the steps of:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a

compound that specifically binds to the polypeptide of claim 1.

26. A method of screening for a compound that modulates the activity of the polypeptide of

5   claim 1, said method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions

permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound

10   with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in

the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a

compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target

15   polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method

comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions

suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

20   c) comparing the expression of the target polynucleotide in the presence of varying amounts of

the compound and in the absence of the compound.

28. A method for assessing toxicity of a test compound, said method comprising:

a) treating a biological sample containing nucleic acids with the test compound;

25   b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at

least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific

hybridization complex is formed between said probe and a target polynucleotide in the biological

sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim

11 or fragment thereof;

30   c) quantifying the amount of hybridization complex; and

d) comparing the amount of hybridization complex in the treated biological sample with the

amount of hybridization complex in an untreated biological sample, wherein a difference in the

amount of hybridization complex in the treated biological sample is indicative of toxicity of the test

compound.

29. A diagnostic test for a condition or disease associated with the expression of XMES in a biological sample comprising the steps of:

a) combining the biological sample with an antibody of claim 10, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex; and

b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

30. The antibody of claim 10, wherein the antibody is:

a) a chimeric antibody,

b) a single chain antibody,

c) a Fab fragment,

d) a F(ab')$_2$ fragment, or

e) a humanized antibody.

31. A composition comprising an antibody of claim 10 and an acceptable excipient.

32. A method of diagnosing a condition or disease associated with the expression of XMES in a subject, comprising administering to said subject an effective amount of the composition of claim 31.

33. A composition of claim 31, wherein the antibody is labeled.

34. A method of diagnosing a condition or disease associated with the expression of XMES in a subject, comprising administering to said subject an effective amount of the composition of claim 33.

35. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 10 comprising:

a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, or an immunogenic fragment thereof, under conditions to elicit an antibody response;

b) isolating antibodies from said animal; and

c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9.

5   36. An antibody produced by a method of claim 35.

37. A composition comprising the antibody of claim 36 and a suitable carrier.

38. A method of making a monoclonal antibody with the specificity of the antibody of claim
10   10 comprising:

a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, or an immunogenic fragment thereof, under conditions to elicit an antibody response;

b) isolating antibody producing cells from the animal;

15   c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells;

d) culturing the hybridoma cells; and

e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9.

20

39. A monoclonal antibody produced by a method of claim 38.

40. A composition comprising the antibody of claim 39 and a suitable carrier.

25   41. The antibody of claim 10, wherein the antibody is produced by screening a Fab expression library.

42. The antibody of claim 10, wherein the antibody is produced by screening a recombinant immunoglobulin library.

30

43. A method for detecting a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9 in a sample, comprising the steps of:

a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9 in the sample.

5   44. A method of purifying a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9 from a sample, the method comprising:

a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and

b) separating the antibody from the sample and obtaining the purified polypeptide having an
10   amino acid sequence selected from the group consisting of SEQ ID NO:1-9.

45. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

46. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

15

47. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

48. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

20   49. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

50. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

51. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

25

52. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

53. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

30   54. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:10.

55. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:11.

56. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:12.

57. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
5   NO:13.

58. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:14.

10   59. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:15.

60. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:16.
15

61. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:17.

62. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
20   NO:18.

<110> INCYTE GENOMICS, INC.
      LAL, Preeti
      YUE, Henry
      HE, Ann
      NGUYEN, Danniel B.
      WALIA, Narinder
      GANDHI, Ameena R.
      AZIMZAI, Yalda
      BANDMAN, Olga
      TANG; Y. Tom
      LU, Yan
      BAUGHN, Mariah R.
      DUGGAN, Brendan M.
      LEE, Sally
      HAFALIA, April
      POLICKY, Jennifer L.

<120> EXTRACELLULAR MESSENGERS

<130> PF-0780 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/210,233; 60/213,465; 60/249,019
<151> 2000-06-06; 2000-06-23; 2000-11-14

<160> 18

<170> PERL Program

<210> 1
<211> 304
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1657368CD1

<400> 1
Met Leu Pro Pro Pro Arg Pro Ala Ala Ala Leu Ala Leu Pro Val
 1   5   10   15
Leu Leu Leu Leu Leu Val Val Leu Thr Pro Pro Pro Thr Gly Ala
                20   25   30
Arg Pro Ser Pro Gly Pro Asp Tyr Leu Arg Arg Gly Trp Met Arg
                35   40   45
Leu Leu Ala Glu Gly Glu Gly Cys Ala Pro Cys Arg Pro Glu Glu
                50   55   60
Cys Ala Ala Pro Arg Gly Cys Leu Ala Gly Arg Val Arg Asp Ala
                65   70   75
Cys Gly Cys Cys Trp Glu Cys Ala Asn Leu Glu Gly Gln Leu Cys
                80   85   90
Asp Leu Asp Pro Ser Ala His Phe Tyr Gly His Cys Gly Glu Gln
                95   100   105
Leu Glu Cys Arg Leu Asp Thr Gly Gly Asp Leu Ser Arg Gly Glu
                110   115   120
Val Pro Glu Pro Leu Cys Ala Cys Arg Ser Gln Ser Pro Leu Cys
                125   130   135
Gly Ser Asp Gly His Thr Tyr Ser Gln Ile Cys Arg Leu Gln Glu
                140   145   150
Ala Ala Arg Ala Arg Pro Asp Ala Asn Leu Thr Val Ala His Pro
                155   160   165
Gly Pro Cys Glu Ser Gly Pro Gln Ile Val Ser His Pro Tyr Asp
                170   175   180
Thr Trp Asn Val Thr Gly Gln Asp Val Ile Phe Gly Cys Glu Val
                185   190   195
Phe Ala Tyr Pro Met Ala Ser Ile Glu Trp Arg Lys Asp Gly Leu

```
                    200                        205                     210
Asp Ile Gln Leu Pro Gly Asp Asp Pro His Ile Ser Val Gln Phe
                    215                        220                     225
Arg Gly Gly Pro Gln Arg Phe Glu Val Thr Gly Trp Leu Gln Ile
                    230                        235                     240
Gln Ala Val Arg Pro Ser Asp Glu Gly Thr Tyr Arg Cys Leu Gly
                    245                        250                     255
Arg Asn Ala Leu Gly Gln Val Glu Ala Pro Ala Ser Leu Thr Val
                    260                        265                     270
Leu Thr Pro Asp Gln Leu Asn Ser Thr Gly Ile Pro Gln Leu Arg
                    275                        280                     285
Ser Leu Asn Leu Val Pro Glu Glu Glu Ala Glu Ser Glu Glu Asn
                    290                        295                     300
Asp Asp Tyr Tyr
```

```
<210> 2
<211> 1438
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4028972CD1

<400> 2
Met Ser Ser Thr Leu His Ser Val Phe Phe Thr Leu Lys Val Ser
  1               5                       10                      15
Ile Leu Leu Gly Ser Leu Leu Gly Leu Cys Leu Gly Leu Glu Phe
                   20                      25                      30
Met Gly Leu Pro Asn Gln Trp Ala Arg Tyr Leu Arg Trp Asp Ala
                   35                      40                      45
Ser Thr Arg Ser Asp Leu Ser Phe Gln Phe Lys Thr Asn Val Ser
                   50                      55                      60
Thr Gly Leu Leu Leu Tyr Leu Asp Asp Gly Gly Val Cys Asp Phe
                   65                      70                      75
Leu Cys Leu Ser Leu Val Asp Gly Arg Val Gln Leu Arg Phe Ser
                   80                      85                      90
Met Asp Cys Ala Glu Thr Ala Val Leu Ser Asn Lys Gln Val Asn
                   95                      100                     105
Asp Ser Ser Trp His Phe Leu Met Val Ser Arg Asp Arg Leu Arg
                   110                     115                     120
Thr Val Leu Met Leu Asp Gly Glu Gly Gln Ser Gly Glu Leu Gln
                   125                     130                     135
Pro Gln Arg Pro Tyr Met Asp Val Val Ser Asp Leu Phe Leu Gly
                   140                     145                     150
Gly Val Pro Thr Asp Ile Arg Pro Ser Ala Leu Thr Leu Asp Gly
                   155                     160                     165
Val Gln Ala Met Pro Gly Phe Lys Gly Leu Ile Leu Asp Leu Lys
                   170                     175                     180
Tyr Gly Asn Ser Glu Pro Arg Leu Leu Gly Ser Arg Gly Val Gln
                   185                     190                     195
Met Asp Ala Glu Gly Pro Cys Gly Glu Arg Pro Cys Glu Asn Gly
                   200                     205                     210
Gly Ile Cys Phe Leu Leu Asp Gly His Pro Thr Cys Asp Cys Ser
                   215                     220                     225
Thr Thr Gly Tyr Gly Gly Lys Leu Cys Ser Glu Asp Val Ser Gln
                   230                     235                     240
Asp Pro Gly Leu Ser His Leu Met Met Ser Glu Gln Gly Arg Ser
                   245                     250                     255
Lys Ala Arg Glu Glu Asn Val Ala Thr Phe Arg Gly Ser Glu Tyr
                   260                     265                     270
Leu Cys Tyr Asp Leu Ser Gln Asn Pro Ile Gln Ser Ser Ser Asp
                   275                     280                     285
Glu Ile Thr Leu Ser Phe Lys Thr Trp Gln Arg Asn Gly Leu Ile
                   290                     295                     300
Leu His Thr Gly Lys Ser Ala Asp Tyr Val Asn Leu Ala Leu Lys
```

```
                      305                    310                    315
Asp Gly Ala Val Ser Leu Val Ile Asn Leu Gly Ser Gly Ala Phe
                      320                    325                    330
Glu Ala Ile Val Glu Pro Val Asn Gly Lys Phe Asn Asp Asn Ala
                      335                    340                    345
Trp His Asp Val Lys Val Thr Arg Asn Leu Arg Gln Val Thr Ile
                      350                    355                    360
Ser Val Asp Gly Ile Leu Thr Thr Thr Gly Tyr Thr Gln Glu Asp
                      365                    370                    375
Tyr Thr Met Leu Gly Ser Asp Asp Phe Phe Tyr Val Gly Gly Ser
                      380                    385                    390
Pro Ser Thr Ala Asp Leu Pro Gly Ser Pro Val Ser Asn Asn Phe
                      395                    400                    405
Met Gly Cys Leu Lys Glu Val Val Tyr Lys Asn Asn Asp Ile Arg
                      410                    415                    420
Leu Glu Leu Ser Arg Leu Ala Arg Ile Ala Asp Thr Lys Met Lys
                      425                    430                    435
Ile Tyr Gly Glu Val Val Phe Lys Cys Glu Asn Val Ala Thr Leu
                      440                    445                    450
Asp Pro Ile Asn Phe Glu Thr Pro Glu Ala Tyr Ile Ser Leu Pro
                      455                    460                    465
Lys Trp Asn Thr Lys Arg Met Gly Ser Ile Ser Phe Asp Phe Arg
                      470                    475                    480
Thr Thr Glu Pro Asn Gly Leu Ile Leu Phe Thr His Gly Lys Pro
                      485                    490                    495
Gln Glu Arg Lys Asp Ala Arg Ser Gln Lys Asn Thr Lys Val Asp
                      500                    505                    510
Phe Phe Ala Val Glu Leu Leu Asp Gly Asn Leu Tyr Leu Leu Leu
                      515                    520                    525
Asp Met Gly Ser Gly Thr Ile Lys Val Lys Ala Thr Gln Lys Lys
                      530                    535                    540
Ala Asn Asp Gly Glu Trp Tyr His Val Asp Ile Gln Arg Asp Gly
                      545                    550                    555
Arg Ser Gly Thr Ile Ser Val Asn Ser Arg Arg Thr Pro Phe Thr
                      560                    565                    570
Ala Ser Gly Glu Ser Glu Ile Leu Asp Leu Glu Gly Asp Met Tyr
                      575                    580                    585
Leu Gly Gly Leu Pro Glu Asn Arg Ala Gly Leu Ile Leu Pro Thr
                      590                    595                    600
Glu Leu Trp Thr Ala Met Leu Asn Tyr Gly Tyr Val Gly Cys Ile
                      605                    610                    615
Arg Asp Leu Phe Ile Asp Gly Arg Ser Lys Asn Ile Arg Gln Leu
                      620                    625                    630
Ala Glu Met Gln Asn Ala Ala Gly Val Lys Ser Ser Cys Ser Arg
                      635                    640                    645
Met Ser Ala Lys Gln Cys Asp Ser Tyr Pro Cys Lys Asn Asn Ala
                      650                    655                    660
Val Cys Lys Asp Gly Trp Asn Arg Phe Ile Cys Asp Cys Thr Gly
                      665                    670                    675
Thr Gly Tyr Trp Gly Arg Thr Cys Glu Arg Glu Ala Ser Ile Leu
                      680                    685                    690
Ser Tyr Asp Gly Ser Met Tyr Met Lys Ile Ile Met Pro Met Val
                      695                    700                    705
Met His Thr Glu Ala Glu Asp Val Ser Phe Arg Phe Met Ser Gln
                      710                    715                    720
Arg Ala Tyr Gly Leu Leu Val Ala Thr Thr Ser Arg Asp Ser Ala
                      725                    730                    735
Asp Thr Leu Arg Leu Glu Leu Asp Gly Gly Arg Val Lys Leu Met
                      740                    745                    750
Val Asn Leu Asp Cys Ile Arg Ile Asn Cys Asn Ser Ser Lys Gly
                      755                    760                    765
Pro Glu Thr Leu Tyr Ala Gly Gln Lys Leu Asn Asp Asn Glu Trp
                      770                    775                    780
His Thr Val Arg Val Val Arg Arg Gly Lys Ser Leu Lys Leu Thr
                      785                    790                    795
Val Asp Asp Asp Val Ala Glu Gly Thr Met Val Gly Asp His Thr
                      800                    805                    810
```

```
Arg Leu Glu Phe His Asn Ile Glu Thr Gly Ile Met Thr Glu Lys
            815                     820                 825
Arg Tyr Ile Ser Val Val Pro Ser Ser Phe Ile Gly His Leu Gln
            830                     835                 840
Ser Leu Met Phe Asn Gly Leu Leu Tyr Ile Asp Leu Cys Lys Asn
            845                     850                 855
Gly Asp Ile Asp Tyr Cys Glu Leu Lys Ala Arg Phe Gly Leu Arg
            860                     865                 870
Asn Ile Ile Ala Asp Pro Val Thr Phe Lys Thr Lys Ser Ser Tyr
            875                     880                 885
Leu Ser Leu Ala Thr Leu Gln Ala Tyr Thr Ser Met His Leu Phe
            890                     895                 900
Phe Gln Phe Lys Thr Thr Ser Pro Asp Gly Phe Ile Leu Phe Asn
            905                     910                 915
Ser Gly Asp Gly Asn Asp Phe Ile Ala Val Glu Leu Val Lys Gly
            920                     925                 930
Tyr Ile His Tyr Val Phe Asp Leu Gly Asn Gly Pro Asn Val Ile
            935                     940                 945
Lys Gly Asn Ser Asp Arg Pro Leu Asn Asp Asn Gln Trp His Asn
            950                     955                 960
Val Val Ile Thr Arg Asp Asn Ser Asn Thr His Ser Leu Lys Val
            965                     970                 975
Asp Thr Lys Val Val Thr Gln Val Ile Asn Gly Ala Lys Asn Leu
            980                     985                 990
Asp Leu Lys Gly Asp Leu Tyr Met Ala Gly Leu Ala Gln Gly Met
            995                    1000                1005
Tyr Ser Asn Leu Pro Lys Leu Val Ala Ser Arg Asp Gly Phe Gln
           1010                    1015                1020
Gly Cys Leu Ala Ser Val Asp Leu Asn Gly Arg Leu Pro Asp Leu
           1025                    1030                1035
Ile Asn Asp Ala Leu His Arg Ser Gly Gln Ile Glu Arg Gly Cys
           1040                    1045                1050
Glu Gly Pro Ser Thr Thr Cys Gln Glu Asp Ser Cys Ala Asn Gln
           1055                    1060                1065
Gly Val Cys Met Gln Gln Trp Glu Gly Phe Thr Cys Asp Cys Ser
           1070                    1075                1080
Met Thr Ser Tyr Ser Gly Asn Gln Cys Asn Asp Pro Gly Ala Thr
           1085                    1090                1095
Tyr Ile Phe Gly Lys Ser Gly Gly Leu Ile Leu Tyr Thr Trp Pro
           1100                    1105                1110
Ala Asn Asp Arg Pro Ser Thr Arg Ser Asp Arg Leu Ala Val Gly
           1115                    1120                1125
Phe Ser Thr Thr Val Lys Asp Gly Ile Leu Val Arg Ile Asp Ser
           1130                    1135                1140
Ala Pro Gly Leu Gly Asp Phe Leu Gln Leu His Ile Glu Gln Gly
           1145                    1150                1155
Lys Ile Gly Val Val Phe Asn Ile Gly Thr Val Asp Ile Ser Ile
           1160                    1165                1170
Lys Glu Glu Arg Thr Pro Val Asn Asp Gly Lys Tyr His Val Val
           1175                    1180                1185
Arg Phe Thr Arg Asn Gly Gly Asn Ala Thr Leu Gln Val Asp Asn
           1190                    1195                1200
Trp Pro Val Asn Glu His Tyr Pro Thr Gly Arg Gln Leu Thr Ile
           1205                    1210                1215
Phe Asn Thr Gln Ala Gln Ile Ala Ile Gly Gly Lys Asp Lys Gly
           1220                    1225                1230
Arg Leu Phe Gln Gly Gln Leu Ser Gly Leu Tyr Tyr Asp Gly Leu
           1235                    1240                1245
Lys Val Leu Asn Met Ala Ala Glu Asn Asn Pro Asn Ile Lys Ile
           1250                    1255                1260
Asn Gly Ser Val Arg Leu Val Gly Glu Val Pro Ser Ile Leu Gly
           1265                    1270                1275
Thr Thr Gln Thr Thr Ser Met Pro Pro Glu Met Ser Thr Thr Val
           1280                    1285                1290
Met Glu Thr Thr Thr Thr Met Ala Thr Thr Thr Thr Arg Lys Asn
           1295                    1300                1305
Arg Ser Thr Ala Ser Ile Gln Pro Thr Ser Asp Asp Leu Val Ser
```

```
                  1310                    1315                    1320
Ser Ala Glu Cys Ser Ser Asp Asp Glu Asp Phe Val Glu Cys Glu
                  1325                    1330                    1335
Pro Ser Thr Ala Asn Pro Thr Glu Pro Gly Ile Arg Arg Val Pro
                  1340                    1345                    1350
Gly Ala Ser Glu Val Ile Arg Glu Ser Ser Ser Thr Thr Gly Met
                  1355                    1360                    1365
Val Val Gly Ile Val Ala Ala Ala Ala Leu Cys Ile Leu Ile Leu
                  1370                    1375                    1380
Leu Tyr Ala Met Tyr Lys Tyr Arg Asn Arg Asp Glu Gly Ser Tyr
                  1385                    1390                    1395
Gln Val Asp Glu Thr Arg Asn Tyr Ile Ser Asn Ser Ala Gln Ser
                  1400                    1405                    1410
Asn Gly Thr Leu Met Lys Glu Lys Gln Gln Ser Ser Lys Ser Gly
                  1415                    1420                    1425
His Lys Lys Gln Lys Asn Lys Asp Arg Glu Tyr Tyr Val
                  1430                    1435
```

<210> 3
<211> 208
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5398353CD1

<400> 3
```
Met Asp Ser Asp Glu Thr Gly Phe Glu His Ser Gly Leu Trp Val
  1               5                   10                  15
Ser Val Leu Ala Gly Leu Leu Gly Ala Cys Gln Ala His Pro Ile
                  20                  25                  30
Pro Asp Ser Ser Pro Leu Leu Gln Phe Gly Gly Gln Val Arg Gln
                  35                  40                  45
Arg Tyr Leu Tyr Thr Asp Asp Ala Gln Gln Thr Glu Ala His Leu
                  50                  55                  60
Glu Ile Arg Glu Asp Gly Thr Val Gly Gly Ala Ala Asp Gln Ser
                  65                  70                  75
Pro Glu Ser Leu Leu Gln Leu Lys Ala Leu Lys Pro Gly Val Ile
                  80                  85                  90
Gln Ile Leu Gly Val Lys Thr Ser Arg Phe Leu Cys Gln Arg Pro
                  95                  100                 105
Asp Gly Ala Leu Tyr Gly Ser Leu His Phe Asp Pro Glu Ala Cys
                  110                 115                 120
Ser Phe Arg Glu Leu Leu Leu Glu Asp Gly Tyr Asn Val Tyr Gln
                  125                 130                 135
Ser Glu Ala His Gly Leu Pro Leu His Leu Pro Gly Asn Lys Ser
                  140                 145                 150
Pro His Arg Asp Pro Ala Pro Arg Gly Pro Ala Arg Phe Leu Pro
                  155                 160                 165
Leu Pro Gly Leu Pro Pro Ala Leu Pro Glu Pro Pro Gly Ile Leu
                  170                 175                 180
Ala Pro Gln Pro Pro Asp Val Gly Ser Ser Asp Pro Leu Ser Met
                  185                 190                 195
Val Gly Pro Ser Gln Gly Arg Ser Pro Ser Tyr Ala Ser
                  200                 205
```

<210> 4
<211> 159
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 71234118CD1

<400> 4

```
Met Ser Phe Phe Asp Asn Ser Thr Ala Ser Ser Asp Asn Trp Lys
1               5                   10                  15
Cys Leu Ser Ser Ala Thr Val Cys Thr Leu Phe Leu Phe Ile Ala
            20                  25                  30
Glu Gln Ser Thr Gly Phe Asn Leu Asp Phe Glu Val Ser Gly Ile
            35                  40                  45
Tyr Gly Tyr Val Met Leu Asp Gly Met Leu Pro Ser Leu His Ala
            50                  55                  60
Leu Thr Cys Thr Phe Trp Met Lys Ser Ser Asp Asp Met Asn Tyr
            65                  70                  75
Gly Thr Pro Ile Ser Tyr Ala Val Asp Asn Gly Ser Asp Asn Thr
            80                  85                  90
Leu Leu Leu Thr Asp Tyr Asn Gly Trp Val Leu Tyr Val Asn Gly
            95                  100                 105
Arg Glu Lys Ile Thr Asn Cys Pro Ser Val Asn Asp Gly Arg Trp
            110                 115                 120
His His Ile Ala Ile Thr Trp Thr Ser Ala Asn Gly Ile Trp Lys
            125                 130                 135
Val Tyr Ile Asp Gly Lys Leu Ser Asp Gly Gly Ala Gly Leu Ser
            140                 145                 150
Val Gly Leu Pro Ile Pro Gly Met Phe
            155

<210> 5
<211> 500
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 240168CD1

<400> 5
Met Ala Ser Val Ala Gln Glu Ser Ala Gly Ser Gln Arg Arg Leu
1               5                   10                  15
Pro Pro Arg His Gly Ala Leu Arg Gly Leu Leu Leu Leu Cys Leu
            20                  25                  30
Trp Leu Pro Ser Gly Arg Ala Ala Leu Pro Pro Ala Ala Pro Leu
            35                  40                  45
Ser Glu Leu His Ala Gln Leu Ser Gly Val Glu Gln Leu Leu Glu
            50                  55                  60
Glu Phe Arg Arg Gln Leu Gln Gln Glu Arg Pro Gln Glu Glu Leu
            65                  70                  75
Glu Leu Glu Leu Arg Ala Gly Gly Gly Pro Gln Glu Asp Cys Pro
            80                  85                  90
Gly Pro Gly Ser Gly Gly Tyr Ser Ala Met Pro Asp Ala Ile Ile
            95                  100                 105
Arg Thr Lys Asp Ser Leu Ala Ala Gly Ala Ser Phe Leu Arg Ala
            110                 115                 120
Pro Ala Ala Val Arg Gly Trp Arg Gln Cys Val Ala Ala Cys Cys
            125                 130                 135
Ser Glu Pro Arg Cys Ser Val Ala Val Val Glu Leu Pro Arg Arg
            140                 145                 150
Pro Ala Pro Pro Ala Ala Val Leu Gly Cys Tyr Leu Phe Asn Cys
            155                 160                 165
Thr Ala Arg Gly Arg Asn Val Cys Lys Phe Ala Leu His Ser Gly
            170                 175                 180
Tyr Ser Ser Tyr Ser Leu Ser Arg Ala Pro Asp Gly Ala Ala Leu
            185                 190                 195
Ala Thr Ala Arg Ala Ser Pro Arg Gln Glu Lys Asp Ala Pro Pro
            200                 205                 210
Leu Ser Lys Ala Gly Gln Asp Val Val Leu His Leu Pro Thr Asp
            215                 220                 225
Gly Val Val Leu Asp Gly Arg Glu Ser Thr Asp Asp His Ala Ile
            230                 235                 240
Val Gln Tyr Glu Trp Ala Leu Leu Gln Gly Asp Pro Ser Val Asp
            245                 250                 255
```

```
Met Lys Val Pro Gln Ser Gly Thr Leu Lys Leu Ser His Leu Gln
              260                   265                   270
Glu Gly Thr Tyr Thr Phe Gln Leu Thr Val Thr Asp Thr Ala Gly
              275                   280                   285
Gln Arg Ser Ser Asp Asn Val Ser Val Thr Val Leu Arg Ala Ala
              290                   295                   300
Tyr Ser Thr Gly Gly Cys Leu His Thr Cys Ser Arg Tyr His Phe
              305                   310                   315
Phe Cys Asp Asp Gly Cys Cys Ile Asp Ile Thr Leu Ala Cys Asp
              320                   325                   330
Gly Val Gln Gln Cys Pro Asp Gly Ser Asp Glu Asp Phe Cys Gln
              335                   340                   345
Asn Leu Gly Leu Asp Arg Lys Met Val Thr His Thr Ala Ala Ser
              350                   355                   360
Pro Ala Leu Pro Arg Thr Thr Gly Pro Ser Glu Asp Ala Gly Gly
              365                   370                   375
Asp Ser Leu Val Glu Lys Ser Gln Lys Ala Thr Ala Pro Asn Lys
              380                   385                   390
Pro Pro Ala Leu Ser Asn Thr Glu Lys Arg Asn His Ser Ala Phe
              395                   400                   405
Trp Gly Pro Glu Ser Gln Ile Ile Pro Val Met Pro Asp Ser Ser
              410                   415                   420
Ser Ser Gly Lys Asn Arg Lys Glu Glu Ser Tyr Ile Phe Glu Ser
              425                   430                   435
Lys Gly Asp Gly Gly Gly Gly Glu His Pro Ala Pro Glu Thr Gly
              440                   445                   450
Ala Val Leu Pro Leu Ala Leu Gly Leu Ala Ile Thr Ala Leu Leu
              455                   460                   465
Leu Leu Met Val Ala Cys Arg Leu Arg Leu Val Lys Gln Lys Leu
              470                   475                   480
Lys Lys Ala Arg Pro Ile Thr Ser Glu Glu Ser Asp Tyr Leu Ile
              485                   490                   495
Asn Gly Met Tyr Leu
              500

<210> 6
<211> 455
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7481107CD1

<400> 6
Met Asp Thr Pro Arg Val Leu Leu Ser Ala Val Phe Leu Ile Ser
  1               5                  10                   15
Phe Leu Trp Asp Leu Pro Gly Phe Gln Gln Ala Ser Ile Ser Ser
              20                   25                    30
Ser Ser Ser Ser Ala Glu Leu Gly Ser Thr Lys Gly Met Arg Ser
              35                   40                    45
Arg Lys Glu Gly Lys Met Gln Arg Ala Pro Arg Asp Ser Asp Ala
              50                   55                    60
Gly Arg Glu Gly Gln Glu Pro Gln Pro Arg Pro Gln Asp Glu Pro
              65                   70                    75
Arg Ala Gln Gln Pro Arg Ala Gln Glu Pro Pro Gly Arg Gly Pro
              80                   85                    90
Arg Val Val Pro His Glu Tyr Met Leu Ser Ile Tyr Arg Thr Tyr
              95                   100                  105
Ser Ile Ala Glu Lys Leu Gly Ile Asn Ala Ser Phe Phe Gln Ser
              110                  115                  120
Ser Lys Ser Ala Asn Thr Ile Thr Ser Phe Val Asp Arg Gly Leu
              125                  130                  135
Asp Asp Leu Ser His Thr Pro Leu Arg Arg Gln Lys Tyr Leu Phe
              140                  145                  150
Asp Val Ser Met Leu Ser Asp Lys Glu Glu Leu Val Gly Ala Glu
              155                  160                  165
```

```
Leu Arg Leu Phe Arg Gln Ala Pro Ser Ala Pro Trp Gly Pro Pro
            170                 175                 180
Ala Gly Pro Leu His Val Gln Leu Phe Pro Cys Leu Ser Pro Leu
            185                 190                 195
Leu Leu Asp Ala Arg Thr Leu Asp Pro Gln Gly Ala Pro Pro Ala
            200                 205                 210
Gly Trp Glu Val Phe Asp Val Trp Gln Gly Leu Arg His Gln Pro
            215                 220                 225
Trp Lys Gln Leu Cys Leu Glu Leu Arg Ala Ala Trp Gly Glu Leu
            230                 235                 240
Asp Ala Gly Glu Ala Glu Ala Arg Ala Arg Gly Pro Gln Gln Pro
            245                 250                 255
Pro Pro Pro Asp Leu Arg Ser Leu Gly Phe Gly Arg Arg Val Arg
            260                 265                 270
Pro Pro Gln Glu Arg Ala Leu Leu Val Val Phe Thr Arg Ser Gln
            275                 280                 285
Arg Lys Asn Leu Phe Ala Glu Met Arg Glu Gln Leu Gly Ser Ala
            290                 295                 300
Glu Ala Ala Gly Pro Gly Ala Gly Ala Glu Gly Ser Trp Pro Pro
            305                 310                 315
Pro Ser Gly Ala Pro Asp Ala Arg Pro Trp Leu Pro Ser Pro Gly
            320                 325                 330
Arg Arg Arg Arg Arg Thr Ala Phe Ala Ser Arg His Gly Lys Arg
            335                 340                 345
His Gly Lys Lys Ser Arg Leu Arg Cys Ser Lys Lys Pro Leu His
            350                 355                 360
Val Asn Phe Lys Glu Leu Gly Trp Asp Asp Trp Ile Ile Ala Pro
            365                 370                 375
Leu Glu Tyr Glu Ala Tyr His Cys Glu Gly Val Cys Asp Phe Pro
            380                 385                 390
Leu Arg Ser His Leu Glu Pro Thr Asn His Ala Ile Ile Gln Thr
            395                 400                 405
Leu Met Asn Ser Met Asp Pro Gly Ser Thr Pro Pro Ser Cys Cys
            410                 415                 420
Val Pro Thr Lys Leu Thr Pro Ile Ser Ile Leu Tyr Ile Asp Ala
            425                 430                 435
Gly Asn Asn Val Val Tyr Lys Gln Tyr Glu Asp Met Val Val Glu
            440                 445                 450
Ser Cys Gly Cys Arg
            455
```

```
<210> 7
<211> 121
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7476245CD1

<400> 7
Met Ala Arg Arg Ala Gly Gly Ala Arg Met Phe Gly Ser Leu Leu
1               5                   10                  15
Leu Phe Ala Leu Leu Ala Ala Gly Val Ala Pro Leu Ser Trp Asp
            20                  25                  30
Leu Pro Glu Pro Arg Ser Arg Ala Ser Lys Ile Arg Val His Ser
            35                  40                  45
Arg Gly Asn Leu Trp Ala Thr Gly His Phe Met Gly Lys Lys Ser
            50                  55                  60
Leu Glu Pro Ser Ser Pro Ser Pro Leu Gly Thr Ala Pro His Thr
            65                  70                  75
Ser Leu Arg Asp Gln Arg Leu Gln Leu Ser His Asp Leu Leu Gly
            80                  85                  90
Ile Leu Leu Leu Lys Lys Ala Leu Gly Val Ser Leu Ser Arg Pro
            95                  100                 105
Ala Pro Gln Ile Gln Tyr Arg Arg Leu Leu Val Gln Ile Leu Gln
            110                 115                 120
```

8/17

Lys

<210> 8
<211> 55
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5819744CD1

<400> 8
Met Thr Ala Ser Glu Cys Gly Thr Gly Asp Gln Ser Cys Arg Arg
 1   5   10   15
Val Ser Trp Val Val Leu Glu Glu Asp Leu Glu Asp Asp Ala Glu
                20   25   30
Lys Asp Lys Leu Asn Arg Arg Leu Val Val Leu Leu Gln Leu Leu
                35   40   45
Tyr Arg Gly Val Lys Pro Ser Ser Ser Val
                50   55

<210> 9
<211> 545
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5378618CD1

<400> 9
Met Ala Leu Arg Arg Gly Gly Cys Gly Ala Leu Gly Leu Leu Leu
 1   5   10   15
Leu Leu Leu Gly Ala Ala Cys Leu Ile Pro Arg Ser Ala Gln Val
                20   25   30
Arg Arg Leu Ala Arg Cys Pro Ala Thr Cys Ser Cys Thr Lys Glu
                35   40   45
Ser Ile Ile Cys Val Gly Ser Ser Trp Val Pro Arg Ile Val Pro
                50   55   60
Gly Asp Ile Ser Ser Leu Ser Leu Val Asn Gly Thr Phe Ser Glu
                65   70   75
Ile Lys Asp Arg Met Phe Ser His Leu Pro Ser Leu Gln Leu Leu
                80   85   90
Leu Leu Asn Ser Asn Ser Phe Thr Ile Ile Arg Asp Asp Ala Phe
                95   100   105
Ala Gly Leu Phe His Leu Glu Tyr Leu Phe Ile Glu Gly Asn Lys
                110   115   120
Ile Glu Thr Ile Ser Arg Asn Ala Phe Arg Gly Leu Arg Asp Leu
                125   130   135
Thr His Leu Ser Leu Ala Asn Asn His Ile Lys Ala Leu Pro Arg
                140   145   150
Asp Val Phe Ser Asp Leu Asp Ser Leu Ile Glu Leu Asp Leu Arg
                155   160   165
Gly Asn Lys Phe Glu Cys Asp Cys Lys Ala Lys Trp Leu Tyr Leu
                170   175   180
Trp Leu Lys Met Thr Asn Ser Thr Val Ser Asp Val Leu Cys Ile
                185   190   195
Gly Pro Pro Glu Tyr Gln Glu Lys Lys Leu Asn Asp Val Thr Ser
                200   205   210
Phe Asp Tyr Glu Cys Thr Thr Thr Asp Phe Val Val His Gln Thr
                215   220   225
Leu Pro Tyr Gln Ser Val Ser Val Asp Thr Phe Asn Ser Lys Asn
                230   235   240
Asp Val Tyr Val Ala Ile Ala Gln Pro Ser Met Glu Asn Cys Met
                245   250   255
Val Leu Glu Trp Asp His Ile Glu Met Asn Phe Arg Ser Tyr Asp

```
                    260                          265                          270
Asn Ile Thr Gly Gln Ser Ile Val Gly Cys Lys Ala Ile Leu Ile
                    275                          280                          285
Asp Asp Gln Val Phe Val Val Val Ala Gln Leu Phe Gly Gly Ser
                    290                          295                          300
His Ile Tyr Lys Tyr Asp Glu Ser Trp Thr Lys Phe Val Lys Phe
                    305                          310                          315
Gln Asp Ile Glu Val Ser Arg Ile Ser Lys Pro Asn Asp Ile Glu
                    320                          325                          330
Leu Phe Gln Ile Asp Asp Glu Thr Phe Phe Val Ile Ala Asp Ser
                    335                          340                          345
Ser Lys Ala Gly Leu Ser Thr Val Tyr Lys Trp Asn Ser Lys Gly
                    350                          355                          360
Phe Tyr Ser Tyr Gln Ser Leu His Glu Trp Phe Arg Asp Thr Asp
                    365                          370                          375
Ala Glu Phe Val Asp Ile Asp Gly Lys Ser His Leu Ile Leu Ser
                    380                          385                          390
Ser Arg Ser Gln Val Pro Ile Ile Leu Gln Trp Asn Lys Ser Ser
                    395                          400                          405
Lys Lys Phe Val Pro His Gly Asp Ile Pro Asn Met Glu Asp Val
                    410                          415                          420
Leu Ala Val Lys Ser Phe Arg Met Gln Asn Thr Leu Tyr Leu Ser
                    425                          430                          435
Leu Thr Arg Phe Ile Gly Asp Ser Arg Val Met Arg Trp Asn Ser
                    440                          445                          450
Lys Gln Phe Val Glu Ile Gln Ala Leu Pro Ser Arg Gly Ala Met
                    455                          460                          465
Thr Leu Gln Pro Phe Ser Phe Lys Asp Asn His Tyr Leu Ala Leu
                    470                          475                          480
Gly Ser Asp Tyr Thr Phe Ser Gln Ile Tyr Gln Trp Asp Lys Glu
                    485                          490                          495
Lys Gln Leu Phe Lys Lys Phe Lys Glu Ile Tyr Val Gln Ala Pro
                    500                          505                          510
Arg Ser Phe Thr Ala Val Ser Thr Asp Arg Arg Asp Phe Phe Phe
                    515                          520                          525
Ala Ser Ser Phe Lys Gly Lys Thr Lys Ile Phe Glu His Ile Ile
                    530                          535                          540
Val Asp Leu Ser Leu
                    545
```

```
<210> 10
<211> 1374
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1657368CB1

<400> 10
ccaagttcgt gggctctctc agaagtcctc aggacggagc agaggtggcc ggcgggcccg 60
gctgactgcg cctctgcttt ctttccataa ccttttcttt cggactcgaa tcacggctgc 120
tgcgaagggt ctagttccgg acactagggt gcccgaacgc gctgatgccc cgagtgctcg 180
cagggcttcc cgctaaccat gctgccgccg ccgcggcccg cagctgcctt ggcgctgcct 240
gtgctcctgc tactgctggt ggtgctgacg ccgcccccga ccggcgcaag gccatcccca 300
ggcccagatt acctgcggcg cggctggatg cggctgctag cggaggcgcga gggctgcgct 360
ccctgccggc cagaagagtg cgccgcgccg cggggctgcc tggcgggcag ggtgcgcgac 420
gcgtgcggct gctgctggga atgcgccaac ctcgagggcc agctctgcga cctggacccc 480
agtgctcact tctacgggca ctgcggcgag cagcttgagt gccggctgga cacaggcggc 540
gacctgagcc gcggagaggt gccggaacct ctgtgtgcct gtcgttcgca gagtccgctc 600
tgcgggtccg acggtcacac ctactcccag atctgccgcc tgcaggaggc ggcccgcgct 660
cggcccgatg ccaacctcac tgtggcacac ccgggggcccc gcgaatcggg gccccagatc 720
gtgtcacatc catatgacac ttggaatgtg acagggcagg atgtgatctt tggctgtgaa 780
gtgtttgcct accccatggc ctccatcgag tggaggaagg atggcttgga catccagctg 840
ccaggggatg acccccacat ctctgtgcag tttaggggtg accccagag gtttgaggtg 900
actggctggc tgcagatcca ggctgtgcgt cccagtgatg agggcactta ccgctgcctt 960
ggccgcaatg ccctgggtca agtggaggcc cctgctagct tgacagtgct cacacctgac 1020
```

```
cagctgaact ctacaggcat cccccagctg cgatcactaa acctggttcc tgaggaggag 1080
gctgagagtg aagagaatga cgattactac taggtccaga gctctggccc atggggggtgg 1140
gtgagcggct atagtgttca tccctgctct tgaaaagacc tggaaagggg agcagggtcc 1200
cttcatcgac tgctttcatg ctgtcagtag ggatgatcat gggaggccta tttgactcca 1260
aggtagcagt gtggtaggat agagacaaaa gctggaggaa ggtagggaga gaaactgaga 1320
ccaggacccg tggggtacaa aggggccat gcaggagata gcctggccag tagg         1374
```

<210> 11
<211> 4541
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4028972CB1

<400> 11
```
tctgtgtgtg tgctgccttc ctcctgtgtg ctttctgtcc ccccatctct gtcttgtctt 60
tcccacttct attgccaaag ggagagatcc tctccgggct gttccctggc ctgtctgctc 120
ctccgggctc tgtcccagca gcgacaatga gctccacact ccactcggtt ttcttcaccc 180
tgaaggtcag catcctgctg gggtccctgc tggggctctg cctgggcctt gagttcatgg 240
gcctccccaa ccagtgggcc cgctacctcc gctgggatgc cagcacacgc agtgacctga 300
gtttccagtt caagaccaac gtctctacgg ggctgctcct ctacctggat gatggcggcg 360
tctgcgactt cctatgcctc tccctggtgg atggccgcgt tcagctccgc ttcagcatgg 420
actgtgccga gactgccgtg ctgtccaaca agcaggtgaa tgacagcagc tggcacttcc 480
tcatggtgag ccgtgaccgc ctgcgcacgg tgctgatgct tgatggcgag ggccagtctg 540
gggagctgca gccccagcgg ccctacatgg atgtggtcag tgacttgttc cttggtggag 600
tccctactga catacgacct tctgccctga cccttgatgg agttcaggcc atgcccggct 660
tcaagggggtt aattctggat ctcaagtatg gaaactcgga gcctcggctt ctggggagcc 720
ggggtgtcca gatggatgcc gagggaccct gtggtgagcg tccctgtgaa aatggtggga 780
tctgctttct cctggacggc caccccacct gtgactgttc taccactggc tatggtggca 840
agctctgctc agaagatgtc agtcaagatc caggcctctc ccacctcatg atgagtgaac 900
aaggtagaag taaagctcga gaggagaatg tggccacttt ccgaggctca gagtatctgt 960
gctacgacct gtctcagaac ccgatccaga gcagcagtga tgaaatcacc ctctccttta 1020
agacctggca gcgtaacggc ctcatcctgc acacgggcaa gtcggctgac tatgtcaacc 1080
tggctctgaa ggatggtgcg gtctcctgg tcattaacct ggggtccggg gcctttgagg 1140
ccattgtgga gccagtgaat ggaaaattca acgacaacgc ctggcatgat gtcaaagtga 1200
cacgcaacct ccggcaggtg acaatctctg tggatggcat tcttaccacg acgggctaca 1260
ctcaagagga ctataccatg ctgggctcgg acgacttctt ctatgtagga ggaagcccaa 1320
gtaccgctga cttgcctggc tcccctgtca gcaacaactt catgggctgc cttaaagagg 1380
ttgtttataa gaataatgac atccgtctgg agctgtctcg cctggcccgg attgcggaca 1440
ccaagatgaa aatctatggc gaagttgtgt ttaagtgtga gaatgtggcc acactggacc 1500
ccatcaactt tgagaccccca gaggcttaca tcagcttgcc caagtggaac actaaacgta 1560
tgggctccat ctcctttgac ttccgcacca cagagcccaa tggcctgatc ctcttcactc 1620
atggaaagcc ccaagagagg aaggatgctc ggagccagaa gaatacaaaa gtagacttct 1680
ttgccgtgga actcctcgat ggcaacctgt acttgctgct tgacatgggc tctggcacca 1740
tcaaagtgaa agccactcag aagaaagcca atgatgggga atggtaccat gtggacattc 1800
agcgagatgg cagatcaggt actatatcag tgaacagcag gcgcacgcca ttcaccgcca 1860
gtggggagag cgagatcctg gacctggaag gagacatgta cctgggaggg ctgccggaga 1920
accgtgctgg ccttattctc cccaccgagc tgtggactgc catgctcaac tatggctacg 1980
tgggctgcat ccgcgaccta ttcattgatg ggcgcagcaa gaacattcga cagctggcag 2040
agatgcagaa tgctgcgggt gtcaagtcct cctgttcacg gatgagtgcc aagcagtgtg 2100
acagctaccc ctgcaagaat aatgctgtgt gcaaggacgg ctggaaccgc ttcatctgcg 2160
actgcaccgg caccggatac tggggaagaa cctgcgaaag ggaggcatcc atcctgagct 2220
atgatggtag catgtacatg aagatcatca tgcccatggt catgcatact gaggcagagg 2280
atgtgtcctt ccgcttcatg tcccagcgag cttatgggct gctggtggct acgacctcca 2340
gggactctgc cgacaccctg cgtctggagc tggatggggg gcgtgtcaag ctcatggtta 2400
acttagactg tatcaggata aactgtaact ccagcaaagg accagagacc ttgtatgcag 2460
ggcagaagct caatgacaac gagtggcaca ccgttcgggt ggtgcggaga ggaaaaagcc 2520
ttaagttaac cgtggatgat gatgtggctg agggtacaat ggtgggagac cataccgtt 2580
tggagttcca caacattgaa acgggaatca tgactgagaa acgctacatc tccgttgtcc 2640
cctccagctt tattggccat ctgcagagcc tcatgtttaa tggccttctc tacattgact 2700
tgtgcaaaaa tggtgacatt gattattgtg agctgaaggc tcgtttttgga ctgaggaaca 2760
tcatcgctga ccctgtcacc tttaagacca agagcagcta cctgagcctt gccactcttc 2820
aggcttacac ctccatgcac ctcttcttcc agttcaagac cacctcacca gatggcttca 2880
ttctcttcaa tagtggtgat ggcaatgact tcattgcagt cgagcttgtc aaggggtata 2940
tacactacgt ttttgacctc ggaaacggtc ccaatgtgat caaaggcaac agtgaccgcc 3000
```

```
ccctgaatga caaccagtgg cacaatgtcg tcatcactcg ggacaatagt aacactcata 3060
gcctgaaagt ggacaccaaa gtggtcactc aggttatcaa tggtgccaaa aatctggatt 3120
tgaaaggtga tctctatatg gctggtctgg cccaaggcat gtacagcaac ctcccaaagc 3180
tcgtggcctc tcgagatggc tttcaggget gtctagcatc agtggacttg aatggacgcc 3240
tgccagacct catcaatgat gctcttcatc ggagcggaca gatcgagcgt ggctgtgaag 3300
gacccagtac cacctgccag gaagattcat gtgccaacca gggggtctgc atgcaacaat 3360
gggagggctt cacctgtgat tgttctatga cctcttattc tggaaaccag tgcaatgatc 3420
ctggcgctac gtacatcttt gggaaaagtg gtgggcttat cctctacacc tggccagcca 3480
atgacaggcc cagcacgcgg tctgaccgcc ttgccgtggg cttcagcacc actgtgaagg 3540
atggcatctt ggtccgcatc gacagtgctc caggacttgg tgacttcctc cagcttcaca 3600
tagaacaggg gaaaattgga gttgtcttca acattggcac agttgacatc tccatcaaag 3660
aggagagaac ccctgtaaat gacggcaaat accatgtggt acgcttcacc aggaacggcg 3720
gcaacgccac cctgcaggtg gacaactggc cagtgaatga acattatcct acaggccggc 3780
agttaaccat cttcaacact caggcgcaaa tagccattgg tggaaaggac aaaggacgcc 3840
tcttccaagg ccaactctct gggctctatt atgatggttt gaaagtactg aacatgtcgg 3900
ctgagaacaa ccccaatatt aaaatcaatg gaagtgttcg gctggttgga gaagtcccat 3960
caattttggg aacaacacag acgacctcca tgccaccaga aatgtctact actgtcatgg 4020
aaaccactac tacaatggcg actaccacaa cccgtaagaa tcgctctaca gccagcattc 4080
agccaacatc agatgatctt gtttcatctg ctgaatgttc aagtgatgat gaagactttg 4140
ttgaatgtga gccgagtaca gcaaaccocca cggagccggg aatcagacgg gttccggggg 4200
cctcagaggt gatccgggag tcgagcagca caacaggga't ggtcgtcggc attgtggctg 4260
ctgccgccct ctgcatcttg atcctcctgt acgccatgta caagtacagg aacagggacg 4320
aggggtccta tcaagtggac gagacgcgga actacatcag caactccgcc cagagcaacg 4380
gcacgctcat gaaggagaag cagcagagct cgaagagcgg ccacaagaaa cagaaaaaca 4440
aggacaggga gtattacgtg taaacatgcg aacactgctc acacgcgagt tttcacagtt 4500
atttctatcc acgcctatga atctttggac ggtgagatct c                      4541
```

<210> 12
<211> 1117
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5398353CB1

<400> 12
```
ctctgcctag cacatccccc ctcccacctc ctcatccaca aaatgtatag gtttgcataa 60
aataaggtgg aaaattagac agcagcgaga tcatgaaggg tgttgaatga caccgagtaa 120
atacacttta acctatagaa ttttaaggca aaaagtgagc tatgacgtct gcaagcaagc 180
ggtaagtaaa gtccggaatc cgggttcgag gctgtcagct gaggatccag ccgaaagagg 240
agccaggcac tcaggccacc tgagtctact cacctggaca actggaatct ggcaccaatt 300
ctaaaccact cagcttctcc gagctcacac cccggagatc acctgaggac ccgagccatt 360
gatggactcg gacgagaccg ggttcgagca ctcaggactg tgggtttctg tgctggctgg 420
tctgctggga gcctgccagg cacaccccat ccctgactcc agtcctctcc tgcaattcgg 480
gggccaagtc cggcagcggt acctctacac agatgatgcc cagcagacag aagcccacct 540
ggagatcagg gaggatggga cggtgggggg cgctgctgac cagagccccg aaagtctcct 600
gcagctgaaa gccttgaagc cgggagttat tcaaatcttg ggagtcaaga catccaggtt 660
cctgtgccag cggccagatg gggccctgta tggatcgctc cactttgacc ctgaggcctg 720
cagcttccgg gagctgcttc ttgaggacgg atacaatgtt taccagtccg aagcccacgg 780
cctcccgctg cacctgccag ggaacaagtc cccacaccgg gaccctgcac cccgaggacc 840
agctcgcttc ctgccactac caggcctgcc ccccgcactc ccggagccac ccggaatcct 900
ggcccccag ccccccgatg tgggctcctc ggaccctctg agcatggtgg gaccttccca 960
gggccgaagc cccagctacg cttcctgaag ccagaggctg tttactatga catctcctct 1020
ttatttatta ggttatttat cttatttatt ttttattttt cttacttgag ataataaaga 1080
gttccagagg agaaaaaaaa aaaaaaaaa aaaaaaa                            1117
```

<210> 13
<211> 2460
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 71234118CB1

<220>

<221> unsure
<222> 2443 :
<223> a, t, c, g, or other

<400> 13
cctgtgtgca gctggcttca caggatcaca ctgtgaattg aacatcaatg aatgtcagtc 60
taatccatgt agaaatcagg ccacctgtgt ggatgaatta aattcataca gttgtaaatg 120
tcagccagga tttttcaggca aaaggtgtga aacaggtatg tatcaactca gtgttattaa 180
taacaatact aacaatagta atataataac aataatacca attttgcctg caccaagcag 240
tgtccgtagg ccaaaaacga tgccaagcac tttgtactca tttaatcttc acagccatcc 300
aacaaactag gtactactac tattatcctc agtggcacag atgatgaaat taaagcttaa 360
aaatttaggg gggaaataga gaaaattgaa tttagtagaa ctcttcatgt tctacttatt 420
tataatagaa atgatacttt gataagccgc ttgacttcag aggtgagact ccatggacat 480
gtttgagaat cagcagatca aatctaaata ctttctcagg agactggagt cttttagctg 540
taagacttgt tttattgagt ggtctgtgca ggtttggggc agagctgaat tgctcatgaa 600
ataggaaaa gccagagcct tgagtagggc agcatgttgc ccttgcagca cctcctcctt 660
ctctcctgtg acatcaatta cttcctctct actgagtcat tctcatcagc atacagacat 720
gctttagtat ctcctggctt tagtagcatt tcaatttcac ctctccttcc agctacttcc 780
tgtttctctg cttcctttca ctattgaatg tttttcctat cttatatact cactggttat 840
tttatttaaa agaatgcact atgtattttg taaataatat gtatttaatt ttgaaactgc 900
ttcttagata atcatattca ctgaaatgag cttctttgac aatagcactg cttcctctga 960
caattggaaa tgtctttcaa gtgccacagt ctgcacactt ttcttattca ttgcagaaca 1020
gtctacaggc tttaacctgg attttgaagt ttctggcatc tatggatatg tcatgctaga 1080
tggcatgctc ccatctctcc atgctctaac ctgtaccttc tggatgaaat cctctgacga 1140
catgaactat ggaacaccaa tctcctatgc agttgataac ggcagcgaca ataccttgct 1200
cctgactgat tataacggct gggttctta tgtgaatggc agggaaaaga taacaaactg 1260
tccctcggtg aatgatggca gatggcatca tattgcaatc acttggacaa gtgccaatgg 1320
catctggaaa gtctatatcg atgggaaatt atctgacggt ggtgctggcc tctctgttgg 1380
tttgcccata cctggtatgt tttaacaaaa tgagtataat tctgttggac tctagaggtg 1440
atgctgctta aatggacatc aggtttgcta gcattgttgt tcttggtttc aaacaaagat 1500
taacagctat ctctctcatt tcctactccc tgattatcct aactgtgggt tattagaagc 1560
ttttttccta agtaagctgt ttgcctctct gtgtgagaaa tgcctttgaa gatctacatt 1620
aagactacca agaagcatgc tgaactagca gctggttgct aaactgtctt ctaaggagat 1680
taattactgg tattttttgt ttattttaag gattgtttg ttttgtaatc agtatataag 1740
caggaccact gctatctatc ttttatggcc cttttcaaat agatcttggc ttgtcttcta 1800
aaaaatactt tttaataaat atcccaggag tttagtttg aagaattgca aactaaattt 1860
cattaggtga cccaggttct cagcattgta catggcaata agcttgttct atctggatac 1920
atctatagaa gtatatgcat atgacattat ttttttacttt gctcatttgt ttcaggtata 1980
ctaaaagaaa attaaataga caaaaatgaa ttaaagtatt agaaactgac taatacattc 2040
tagaagtaag ttgtatgaaa ctaggatacg tatactaata aatatagcac aaaaatacta 2100
aacttcataa aaatgcggtg agatgaccct tgcgaagcct caaatacaat acttatttta 2160
tttattagaa attattggcc aggtacggtg gctcacgcct gtaatcccag cactttggaa 2220
ggctgaggag ggtggatcac gaggtcagga gttggagacc gccctggaca atatggtgaa 2280
gccctatctc tactaaaaac gcacagaaaa attagccagg catggtggta gatgcctgta 2340
attccagcta ctcaggaggc tgaggcagga gaatcgcttg aacccaggag gcgaagattg 2400
ccgtgagcgg agggcatgcg actcgaggcc aagaatgggg agnggcaggt agagggacgg 2460

<210> 14
<211> 2601
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 240168CB1

<400> 14
gggttcgggg gcgccgcgct gtgaggccgg ggcctagagc cagccgcggc cgcgcaggag 60
gggcccaggg cccgcgctcg cccgcgtccc cgccttcctc ccgcgctcag ccccgcctcg 120
gctcgctgcc cttggctctc gtcgccatgg cctccgtcgc ccaggagagc gcgggctcgc 180
agcgccggct accgccgcgt cacggggcgc tgcgcgggct gctactgctc tgcctgtggc 240
tgccaagcgg ccgtgcggcc ttgccgcccg cggcgccgct gtccgaactg cacgcgcagc 300
tgtcgggcgt ggagcagctg ctggaggagt tccgccggca actgcagcag gagcggcctc 360
aggaggagct ggagctggag ctgcgcgagg cgcgggccc ccaggaggac tgcccgggcc 420
cgggcagcgg cggctacagc gcaatgcctg acgccatcat ccgcaccaag gactccctgg 480
cggcgggtgc cagcttcctg cgggcgccgg cggccgtgcg gggctggcgg caatgcgtgg 540
cggcctgctg ctccgagccg cgctgctccg tggccgtggt ggagctgccc cggcgccccg 600

```
cgcccccggc agccgtgctc ggctgctacc tcttcaactg cacggcgcgc ggccgcaacg 660
tctgcaagtt cgcgctgcac agcggctaca gcagctacag cctcagccgc gcgccggacg 720
gcgccgccct ggccaccgcg cgcgccctgc cgcccagga aaaggatgcg cctccactta 780
gcaaggctgg gcaggatgtg gttctgcatc tgcccacaga cggggtggtt ctagacggcc 840
gcgagagcac agatgaccac gccatcgtcc agtatgagtg ggcactgctg caggggggacc 900
cgtcagtgga catgaaggtg cctcaatcag gaaccctgaa gctgtcccac ctacaggagg 960
gaacctacac cttccagctg accgtgacgg acactgccgg gcagagaagc tctgacaacg 1020
tgtcagtgac agtgcttcgc gcagcctact ccacaggagg atgtttgcac acttgctcac 1080
gctaccactt cttctgtgac gatggctgct gcattgacat cacgctcgcc tgcgatggag 1140
tgcagcagtg tcctgatggg tctgatgaag acttctgcca gaatctgggc ctggaccgca 1200
agatggtaac ccacacggca gctagtcctg ccctgccaag aaccacaggg ccgagtgaag 1260
atgcagggggg tgactccttg gtggaaaagt ctcagaaagc cactgcccca aacaagccac 1320
ctgcattatc aaacacagag aagaggaatc attccgcctt ttggggacca gagagtcaaa 1380
tcattcctgt gatgccagat agtagttcct cagggaagaa cagaaaagag gaaagttata 1440
tatttgagtc aaagggtgat ggaggaggag gggaacaccc agccccagaa acaggtgcag 1500
tgctacccct ggcgctgggt ttggctatca ctgctctgct gcttctcatg gttgcatgcc 1560
gactacgact ggtgaaacag aaactgaaaa aagctcgtcc cattacatct gaggaatcgg 1620
actacctcat aaatgggatg tatctatagt aatgtaattt caataccttg gggcagggac 1680
atgttttgtt tataatttat acatctatta agttctggat atttacagct tctttttgttt 1740
ttaattgggc cagaagattc tgcaaatccc aaatctttct ttattattta ttgtaaaaaa 1800
agtttcctta gaagtcataa aatattttga aatttagaga ggaattcatg attaaagatt 1860
cctaaaaata taattctgat ttatgtaagc tgtccctgaa aatagaaatg tgtacttagc 1920
tgagagaaaa ttcagcatct caggaggtgg tattaggatg actgtgttaa cccattacct 1980
tttagaagcc aactgttggc cccttaccat gctggactgc tataggccca gcttccccttt 2040
gttctgtggc ccttttcttc ctccttgaag ctcccagtat tcttttttctt ttcccctcta 2100
aacctgtttc tgagagtgga tctcaagcaa gttcatgcct tcaatcagat gttacttagg 2160
gtgggtatac ctaaattata aaccttatgt acaagtcagt aagccttagg gaaggtgagt 2220
gtgggtcctt cctaatccct ctgacgtcat gtcatatagg tggctgcctc cttagactga 2280
cctttgggag aaaaaaaccc cagactttga attagtaaca gctctaagat ggtcatgcag 2340
tgagatagga aatcaagatg gaagcagaga atctggcatg ccaaaaacta acagaaactt 2400
agttgaaggc aaagagagca aggagaacgt ttaatacttc attacatcaa atcaacactg 2460
ctccatggtg agagcacagc aactcattta tatatatata tataggcttt gttgatgaaa 2520
aacaacaatt gaagagagca cgttgagtgg attcctgggt acagcttttg taaaaatgtc 2580
accatggctt tcatccaatg g                                             2601
```

```
<210> 15
<211> 2791
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7481107CB1

<400> 15
cttggggaag gaggaagtcc tgcaggcggg agggaaagaa gagagggaaa atggggatgc 60
agtggaggcg ggggggcaggc cgcgagaggg agaggatccc gggagcagac gaagaagtgg 120
agcagctaaa gtctgcgtca gaagaggttg gggactgcga gaggagaggc tggggcctgc 180
agggggagcgc agcagctttt agcatcgatc caaactctaa agactcgtgg cctttgcctg 240
acctcgaggg tcgggaatag acgcctgtct ttgtggagag cgatacccaa ccgagaaaat 300
ggggctgttc cgagctgggc cctgcgcctg gcccaggggcg aggcttctct ggctccggggc 360
tggccccctga ggggcgcaaa cggcaggcct ggcggttggg gccgaggagg gaaggtacca 420
ccgccccgag gcagcacgca gcctgcagca gaggcgcctg ctccgagctg tctcttggggg 480
gcgccgccgc cgcttccctc ctccgggggcc gctcgctccc aggaaagtgg aggcggctgg 540
cgaggaccga gagccggggc cgcgctgcgg agggaccaca cctccgggag ttcgaggggg 600
accctggcgc ggcgggccag cctttgctcc ccggccacgg gccggcagcg cccgccttcc 660
cccggtcagc gcttgcggcc cgcgccgcgc gcaccgcccg gcaaccccgc gcgcgtcccg 720
cgggggcgct gcgtcttcct gccacaccgg cgcaccgcgg ccctctccc ccacacctcc 780
ggcccgcacc acccggctct cctcccaccc tccccacccc tcctctgccc tccctcccca 840
ttcctccct cccggcgagg ggcgggaggg ggcgtggcgg ggccggggtt tgtgtggctg 900
ggaccccggct cctcgcactc cgagtccgcc cgaggagccg ggccccggcc gctgtccagc 960
cgctccgtgc ccgcgcgtc ctgcgccgcc gccaccgcct cctggggaga cgcagccact 1020
tgcccgccat ggatactccc agggtcctgc tctcggccgt cttcctcatc agttttctgt 1080
gggatttgcc cggtttccag caggcttcca tctcatcctc ctcgtcgtcc gccgagctgg 1140
gttccaccaa gggcatgcga agccgcaagg aaggcaagat gcagcggggcg ccgcgcgaca 1200
gtgacgcggg ccgggagggc caggaaccac agccgcggggc tcaggacgaa ccccgggctc 1260
agcagccccg ggcgcaggag ccgccaggca ggggtccgcg cgtggtgccc cacgagtaca 1320
```

```
tgctgtcaat ctacaggact tactccatcg ctgagaagct gggcatcaat gccagctttt 1380
tccagtcttc caagtcggct aatacgatca ccagctttgt agacagggga ctagacgatc 1440
tctcgcacac tcctctccgg agacagaagt atttgtttga tgtgtccatg ctctcagaca 1500
aagaagagct ggtgggcgcg gagctgcggc tctttcgcca ggcgccctca gcgccctggg 1560
ggccaccagc cgggccgctc cacgtgcagc tcttcccttg cctttcgccc ctactgctgg 1620
acgcgcggac cctggacccg caggggggcgc cgccggccgg ctgggaagtc ttcgacgtgt 1680
ggcagggcct gcgccaccag ccctggaagc agctgtgctt ggagctgcgg gccgcatggg 1740
gcgagctgga cgccggggag gccgaggcgc gcgcgcgggg accccagcaa ccgccgccccc 1800
cggacctgcg gagtctgggc ttcggccgga gggtgcggcc tccccaggag cgggccctgc 1860
tggtggtatt caccagatcc cagcgcaaga acctgttcgc agagatgcgc gagcagctgg 1920
gctcggccga ggctgcgggc ccgggcgcgg gcgccgaggg gtcgtggccg ccgccgtcgg 1980
gcgccccgga tgccaggcct tggctgccct cgcccggccg ccggcggcgg cgcacggcct 2040
tcgccagtcg ccatggcaag cggcacggca agaagtccag gctacgcgctgc agcaagaagc 2100
ccctgcacgt gaacttcaag gagctgggct gggacgactg gattatcgcg ccctggagt 2160
acgaggccta tcactgcgag ggtgtatgcg acttcccgct gcgctcgcac ctggagccca 2220
ccaaccacgc catcatccag acgctgatga actccatgga ccccggctcc accccgccca 2280
gctgctgcgt gcccaccaaa ttgactccca tcagcattct atacatcgac gcgggcaata 2340
atgtggtcta caagcagtac gaggacatgg tggtggagtc gtgcggctgc aggtagcggt 2400
gcctttcccg ccgccttggc ccggaaccaa ggtgggccaa ggtccgcctt gcaggggagg 2460
cctggctgca gagaggcgga ggaggaagct ggcgctgggg gaggctgagg gtgagggaac 2520
agcctggatg tgagagccgg tgggagagaa gggagcgcag acttcccagt aacttctacc 2580
tgccagccca gagggaaata tggattttca caccttgcct ggacaccctg gaaaaacaag 2640
ccaaggagga tttctttagt tctgtttctc tctctctctc tctctctctc tctctctctc 2700
tctctctctc tctctctctc tatcagtgtg tctgtgtaat cccatgtgtg tcatacagct 2760
cgagatataa tgcgccagac ggcacaacct c                              2791
```

<210> 16
<211> 709
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7476245CB1

<400> 16
```
ctcgagccgc gcgcccgaac gaagccgcgg cccgggcaca gccatggccc ggcgggcggg 60
gggcgctcgg atgttcggca gcctcctgct cttcgccctg ctcgctgccg gcgtcgcccc 120
gctcagctgg gatctcccgg agccccgcag ccgagccagc aagatccgag tgcactcgcg 180
aggcaacctc tgggccaccg gtcacttcat gggcaagaag agtctggagc cttccagccc 240
atccccattg gggacagctc cccacacctc cctgagggac cagcgactgc agctgagtca 300
tgatctgctc ggaatcctcc tgctaaagaa ggctctgggc gtgagcctca gccgcccccgc 360
accccaaatc cagtacagga ggctgctggt acaaatactg cagaaatgac accaataatg 420
gggcagacac aacagcgtgg cttagattgt gcccacccag ggaaggtgct gaatgggacc 480
ctgttgatgg ccccatctgg atgtaaatcc tgagctcaaa tctctgttac tccattactg 540
tgatttctgg ctgggtcacc agaaatatcg ctgatgcaga cacagattat gttcctgctg 600
tatttcctgc ttccctgttg aattggtgaa taaaaccttg ctctttacat aaaaaaaaaa 660
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaacaaaaaa aaaaaaaaa          709
```

<210> 17
<211> 753
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5819744CB1

<400> 17
```
ctggacctgg gaccatgtct tgtattgcca gcacagggcc ctgaaacaca agagatattg 60
ggcaagggat agaccagtgg aagagaggta gaaagcaagt ccagagataa agatttagaa 120
aggctgcaga ctaaaatgac tgcttcagaa tgtgggacag gagaccagag ctgcagaaga 180
gtcagctggg tggtattgga ggaggacctg gaagatgacg ctgagaaaga caagcttaac 240
aggagactgg ttgttctctt gcagctgctt tacaggggag tgaagccttc ctcctccgtc 300
tgatgctggg cctgccttcc ctgctgtgca gctactctgt gtctgtgggt ctagtacctt 360
catttctcat tatgttcaca caagtcaggt tgttggggag ctgagctctt acagtcccat 420
gggggaagtg gagagccaaa ggtctcactt tttgtgattc tttggagatt ttatgcaccc 480
```

```
tagaaatttg agttttccct ggtgtcttct gggactggta ctgcataccc agggtcaccc 540
tgtttggggg cttttccaag cggagtttgt gatcataatg agttatattt ttcctctttt 600
ctggataaga acaatatcct tgccattcag tctccaaata tagcatgtgg ttatgataga 660
agaatatcct tatgttagga aagaaattgg cagtctcaat gaggcagaaa tgtacaattt 720
agaaaacatg tataaattac aacagcgggt gta                              753
```

```
<210> 18
<211> 3413
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5378618CB1

<400> 18
cgtctgcagc ggcgccgggt ccgagcgcgc ggcgcggcgg tgggggtcgg ggcccgggcg 60
gggagcgggg accgggcatg gcgctgcgga gaggcggctg cggagcgctc gggctgctgc 120
tgctgctgct gggcgccgcg tgcctgatac cgcggagcgc gcaggtgagg cggctggcgc 180
gctgccccgc cacttgcagc tgtaccaagg agtctatcat ctgcgtgggc tcttcctggg 240
tgcccaggat cgtgccgggc gacatcagct ccctgagcct ggtaaatggg acgtttttcag 300
aaatcaagga ccgaatgttt tcccatctgc cttctctgca gctgctattg ctgaattcta 360
actcattcac gatcatccgg gatgatgctt ttgctggact ttttcatctt gaatacctgt 420
tcattgaagg gaacaaaata gaaaccattt caagaaatgc ctttcgtggc ctccgtgacc 480
tgactcacct ttctttggcc aataaccaca taaaagcact accaagggat gtcttcagtg 540
atttagactc tctgattgaa ctagatttga ggggtaataa atttgaatgt gactgcaaag 600
ccaagtggct atacctgtgg ttgaagatga caaattccac cgtttctgat gtgctgtgta 660
ttggtccacc agagtatcag gaaaagaaagc taaatgacgt gaccagcttt gactatgaat 720
gcacaactac agattttgtt gttcatcaga cttтacccta ccagtcggtt tcagtggata 780
cgttcaactc caagaacgat gtgtacgtgg ccatcgcgca gcccagcatg gagaactgca 840
tggtgctgga gtgggaccac attgaaatga atttccggag ctatgacaac attacaggtc 900
agtccatcgt gggctgtaag gccattctca tcgatgatca ggtctttgtg gtggtagccc 960
agctcttcgg tggctccac atttacaaat acgacgagag ttggaccaaa tttgtcaaat 1020
tccaagacat agaggtctct cgcatttcca agcccaatga catcgagctg tttcagatcg 1080
acgacgagac gttctttgtc atcgcagaca gctcaaaggc tggtctgtcc acagtttata 1140
aatggaacag caaaggattc tattcttacc agtcactgca cgagtggttc agggacacgg 1200
atgcggagtt tgttgatatc gatggaaaat cgcatctcat cctgtccagc cgctcccagg 1260
tccccatcat cctccagtgg aataaaagct ctaagaagtt tgtcccccat ggtgacatcc 1320
ccaacatgga ggacctactg gctgtgaaga gcttccgaat gcaaaaatacc ctctaccttt 1380
cccttacccg cttcatcggg gactcccggg tcatgaggtg gaacagtaag cagtttgtgg 1440
agatccaagc tcttccatcc cggggggcca tgaccctgca gcccttttct tttaaagata 1500
atcactacct ggccctgggg agtgactata cattctctca gatataccag tgggataaag 1560
agaagcagct attcaaaaag tttaaggaga tttacgtgca ggcgcctcgt tcattcacag 1620
ctgtctccac cgacaggaga gatttcttttt ttggcatcacag tttcaaaggg aaaacaaaga 1680
tttttgaaca tataattgtt gacttaagtt tgtgaaggtg tggtgggtga aactaagaga 1740
aatgtagcat tagctctcac aaaagaggac caagaaaaat caacaaacaa atcaaagcca 1800
ggctcagagc tctgaaatta aaaagcactg aaatagttag atgttttcaa actttttagaa 1860
ctcacatttt aatcagggat tgcatttatt ggctaactgc atgacatgcc cattctacca 1920
tttaaaaaaa aatcttaaag cctgtaattt ctgagaaaag agtacagcat ttactcttat 1980
catctagaaa tgtaatatgc ttccccccccg cttttttgatg aggaagagaa caattggata 2040
agatgggaca gcacttataa tgaaataaaa aaaaactttg agcccctctc attccacttt 2100
agcaatcttt ttggtaagaa ctcttaaagc caaagtctg ctgaaaagat ttgctgatta 2160
ttagtttaaa aatcttgtaa cactcagcag tgctatttg agtcatccca gtttcctgaa 2220
agtaatgccc agtcttcctg aatcctcctt aatagcagaa ccttggtgat tttgttggct 2280
catatgaatg cttgtactgg atatgttaac aatttagtgt ttgacattgc ttcctctgcc 2340
acaaagacaa tactctggtg acacatgtct agacccagca caggctgtag gcccaggagt 2400
gactcaaagg agtttttccc tctttcttac ggttcaaagg tgaccctggt ggtggccaga 2460
gcagtaatgc ttgtttgatg ctcttcatgg ctcatctgct tctcagaacc cacccgttga 2520
gtttgtgggt aaccagcagg caggccaaag actggtgctt ttcatttcat cctttagagg 2580
gatgaaacag ttatttccgt ctgatgagca ttcggtagaa ttttttgaagt gagattttat 2640
gaagtcaaag gggacttttac acagatctcg acctgctttg aaacctagag gtggcccttt 2700
gatttgtgcg tgtccttgcc ctctggacaa cttaatattt caagtaatcg aataccaact 2760
tccctgccag cccacctgcc ttccgccccg cttgtgtaac agtcctgttt tgttgagttg 2820
ctgctattgc actgccagtg cagcccacac caaatcacaa cccaagatac tcagatagga 2880
agactccttc ctctcccagt actttaccaa aggaaccccc gccaggaccc acatgggggcc 2940
acgtgttggc agtggaatca gcctgtgcag gctggggatc tcaggctgat cagtaggggc 3000
cagctttgga gccagccaag ctgaatccca cactccaggt ctgtgctcaa gagaccagat 3060
```

```
ggtgtatttc caaatgggcc tctctggtat gggcaatagg caagctcctg gggtctggtt 3120
atgtggaaga ttcttagtgg atgttccgcc tggttagctg gttctcttca gagaatataa 3180
agtgaatgcc tttaggggta gctctgaaag agaaacccaa caacttcatt cctagccatg 3240
aaagtagcac gatcatattg tactgtattg ttattgtaaa atgattattt gccatgtcat 3300
gagtaggtag atgttttgcc acaaatatga aagtgtttgt tgttcctgac tttaagccat 3360
gaagattgag accaataaat agcactcaga ggaatgaaaa aaaaaaaaaa agg        3413
```